155
12.8.7

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

478Z^25

VOLUME 64

1950

Editor
HERBERT S. CONRAD

Federal Security Agency
Office of Education
Washington 25, D.C.

Consulting Editors

DONALD E. BAIER
FRANK A. BEACH
ROBERT G. BERNREUTER
WILLIAM A. BROWNELL
HAROLD E. BURTT
JERRY W. CARTER, JR.
CLYDE H. COOMBS
ETHEL L. CORNELL
JOHN G. DARLEY
JOHN F. DASHIELL
EUGENIA HANFMANN
EDNA HEIDBREDER
HAROLD E. JONES
DONALD W. MACKINNON
LORRIN A. RIGGS
CARL R. ROGERS
SAUL ROSENZWEIG
E. DONALD SISSON
KENNETH W. SPENCE
ROSS STAGNER
PERCIVAL M. SYMONDS
JOSEPH TIFFIN
LEDYARD R TUCKER
JOSEPH ZUBIN

LIBRARY

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION
1515 MASSACHUSETTS AVE. N.W., WASHINGTON 5, D.C.

VOLUME 64
NUMBER 1

WHOLE NO. 307
1950

Psychological Monographs:
General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

25
4782

Patterns of Personality Rigidity and Some of Their Determinants

By

SEYMOUR FISHER

Elgin State Hospital
Elgin, Illinois

Accepted for publication, June 10, 1949

Price $1.00

Published by

THE AMERICAN PSYCHOLOGICAL ASSOCIATION
1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

COPYRIGHT, 1950, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION

ACKNOWLEDGEMENTS

The writer is extremely grateful to Dr. David Shakow for his supervision and persistently helpful criticism of this research over a period of more than a year. Special thanks are due also to Dr. Forrest Kingsbury, Dr. James Miller, Mr. John Butler, and Dr. William Stephenson for criticism and comment at various stages in the development of the project.

The writer is grateful to the following individuals for aid in obtaining subjects: Mr. Leroy Wauck, Chicago State Hospital; Dr. Phyllis Wittman, Elgin State Hospital; Dr. Louis Steinberg, Elgin State Hospital; Dr. Julian Pathman, Illinois Neuropsychiatric Institute; and Dr. Allen Rosenwald, Illinois Neuropsychiatric Institute.

Seymour Fisher

478Z

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

Chapter

I. HYPOTHESIS AND PURPOSE 1

II. SUBJECTS 4

A. Normals 4

B. Hysterics 5

C. Paranoids 5

III. GENERAL PROCEDURE AND MEASUREMENT TECHNIQUES 6

IV. ANALYSIS OF DATA 9

A. Validity Criteria 9

B. Preliminary Analysis of Individual Tasks in the Rigidity Battery 12

C. Analysis of Total Rigidity Score 15

D. Analysis in Terms of Levels of Rigidity 18

E. Analysis of Individual Rigidity Patterns 23

F. Diagnostic Differentiation 34

V. GENERAL DISCUSSION OF DATA 36

VI. GENERAL SUMMARY OF FINDINGS 41

APPENDIXES 43

BIBLIOGRAPHY 48

CHAPTER I

HYPOTHESIS AND PURPOSE

ONE of the most striking deficiencies in previous studies of personality rigidity phenomena revolves about the fact that with but a few limited exceptions they have all confined themselves to measuring rigidity in terms of very simple motor or perceptual factors. No attempt has been made to sample rigidity behavior simultaneously at *different* levels of *complexity* or at different levels of *emotional involvement.* Consequently, there has been little basis for deciding such fundamental questions as to how generalized rigidity may be in the individual personality or whether there are distinctly different kinds of rigidity. This failure has at least in part been due to the fact that no one has formulated a general hypothesis which would define what is common to rigid behavior at different levels. For example, one may easily say that perseverative behavior is a *sign of rigidity* when the performance of some simple motor task is involved; but how does one apply this kind of simple perseveration concept to the detection of rigidity in such complex behavior as an individual's aesthetic preferences or his interpretation of an event with emotional significance? The present study was undertaken to attack some of these problems—to explore in a broader sense the whole question of personality rigidity. In line with the above discussion it was necessary first of all to develop a hypothesis that would define how rigidity might manifest itself in diverse responses to diverse materials. A presentation of this hypothesis and of the specific objectives of the study follow.

The hypothesis that forms the basis for the measurement procedures utilized in this study is that there are persistent personality rigidity trends which are relatively independent of intelligence.[1] It will be assumed that such rigidity trends reveal themselves in the degree to which any given individual is able to indicate in some *behavioral way* his ability to utilize *alternate modes of response* when dealing with problems or situations requiring adjustment. That is, it is postulated that given, for example, two individuals, A and B, of equal formal intelligence, who are faced with the necessity of solving some problem with perhaps six alternative modes of solution, that A will be able to utilize all six of the possible modes of solution, and B will utilize perhaps only two of them, because of personality differentials in rigidity. In these terms an individual's rigidity in expressing aesthetic preferences, or his rigidity in interpreting a picture with emotional meaning, or his rigidity in sorting varied geometric objects into different classes are all measurable on a related continuum—viz., the *number* of his aesthetic preferences, the *number* of his equivalent picture interpretations, and the *number* of his equivalent sortings. This reduction of the measurement of rigidity to *number* of equivalent alternatives utilized is a simplification dictated by the needs of experimental control; but it is also a kind of simplification that is based on a concept that has been implicit in almost all past rigidity measures.

---

[1] That is, intelligence in a limited formal or psychometric sense. It is recognized that intellectual factors pervade most areas of reaction. The assumption here is that intelligence in the formal sense is not significantly related to rigidity behavior resulting from the particular character of the personality structure.

It is important that it be clearly understood to what degree this brief hypothesis really does neglect the complexity of factors involved in rigidity phenomena. One suspects that if sufficiently sensitive measuring instruments were available, it would be possible to analyse rigidity phenomena in many different dimensions. Thus, *hypothetically* one might be able to measure rigidity as it affects *perception* of situations, as it affects *subjective reactions* to situations, and of course as it affects *overt behavioral* reactions to situations. Furthermore, one might be able to describe rigidity in each of these dimensions in terms of a large number of descriptive continua (e.g., quickness with which evoked, degree of persistence after arousal, and degree of generalization). Ideally, it would be well to measure as many of these phases of rigidity as possible. But it has been necessary here to treat the problem in a simpler fashion: to confine postulations to *overt* behavioral manifestations of rigidity and to limit them to rigidity manifestations conceived to exist on a single restricted continuum (viz., number of utilized alternatives). This has been true simply because there are so few known techniques available for getting at rigidity phenomena in any dimension and because the demands of experimental control place a limit on the number of variables one may consider simultaneously.

Of course, the emphasis that has been placed here upon the overt behavioral aspects of rigidity phenomena is not entirely the result of limitations raised by experimental control and available modes of measurement. It seems to be justified to some extent by the fact that in everyday situations we judge others to be rigid or flexible in terms of their overt behavior. The flexible person shows that he is flexible by dealing with situations in a manner that can be described as "flexible." The rigid person actually *behaves* rigidly in his overt response to various stimuli. Indeed, it is questionable whether mere subjective awareness of alternatives in a situation is a significant index of personality flexibility.

This project has the following objectives:

1. To set up a battery of choice or problem situations in which there are *variable numbers of roughly equivalent modes of response,* and then to determine whether individuals show any measurable difference in the degree to which they utilize *few or many alternatives* from *situation to situation.* The range of situations used in the battery is such as to include reactions to materials varying in meaningfulness, complexity, and emotional involvement.

2. To relate any observed individual consistencies in grasping and utilizing alternatives to measures of personality. The personality measures include the Rorschach test and Guilford's[2] STDCR (12).

3. Furthermore, to determine whether observed rigidity consistencies are correlated with different "types" of functional maladjustment. This is attempted by using subjects who are psychiatrically diagnosed as conversion hysterics, and paranoid schizophrenics, and also normal subjects.

Summarizing and restating the above in another form, it may be said that the study aims at answering such questions as the following:

---

[2] A questionnaire which contains 175 questions pertaining to subjective feelings, degree of social participation, mood changes, etc. It presumably measures five personality factors.

1. Are individuals *consistently* rigid? That is, do they show a *constant level* of rigidity that cuts across a wide range of their responses?

2. Are there distinctly *different kinds* of rigidity? If so, what is the relative importance of each in the personality structure?

3. Do individuals who are in general emotionally constricted (as indicated by the Rorschach test) show a corresponding rigidity in other phases of their behavior?

4. Does an individual's manner of describing himself (as indicated by the Guilford STDCR) have any relationship to rigidity?

5. Do neurotics show specific rigidity trends not manifested by normal subjects? In what way is the rigidity of the paranoid schizophrenic similar to, or different from, the rigidity of the conversion hysteric?

CHAPTER II

SUBJECTS

SIXTY subjects participated in the project. They comprised three distinct groups: twenty normals, twenty conversion hysterics, and twenty paranoid schizophrenics. To hold the sex factor constant only women subjects were used. An attempt was made also to keep the intelligence factor relatively constant by selecting only those persons whose intelligence scores fell roughly within the 95 to 110 IQ range. Actually, when final IQ means were computed for each group it was found that the normal subjects exceeded the hysterics by six points and the paranoids by almost five points. There was not a significant difference between the paranoids and hysterics. As will be shown later, the influence of the intellectual factor on the total results was negligible. Each group will now be described in detail.

Normals

The normal group consisted of nine women who were in their third or fourth year of training as occupational therapists; five student nurses in their last year of training; two student social workers in their first year of graduate work; and four individuals employed as stenographers in a hospital setting. The age range was from 20 to 45—with the majority falling in the twenties. The stenographers in the group were obtained when they responded to a posted notice, but all others were selected from larger available groups because they fell within the 95-110 IQ range. No attempt was made to screen the group in advance for possible pathology. It was assumed that the individuals in the group were "healthy" or "normal" because they were not hospitalized and appeared to be functioning at their particular jobs without any gross signs of maladjustment. In general, the participants were all very cooperative during examination and expressed much spontaneous interest toward various of the tasks. At no time was the real purpose of the testing revealed. The various procedures were simply described as a series of tests that were being given to different groups for comparative purposes. With but a few exceptions all testing was completed in one session.

TABLE 1

DISTRIBUTION OF AGE IN YEARS, EDUCATION IN YEARS, INTELLIGENCE, AND OCCUPATION IN THE EXPERIMENTAL GROUPS
($N=60$ Women)

Group	Age		Intelligence		Education		Occupation
	Mean	Range	Mean	S.D.	Mean	Range	
NORMALS	24	20–45	107.2[a]	3.8	15	12–16	9 occupational therapy students 5 student nurses 2 student social workers 4 stenographers
CONVERSION HYSTERICS	35	18–50	100.6[b]	4.9	10	8–12	9 housewives 4 stenographers 7 in miscellaneous factory and clerical work
PARANOID SCHIZOPHRENICS	37	21–55	102.6	5.3	10.8	7–16	10 housewives 5 clerical workers 1 florist 4 miscellaneous factory work

[a] Exceeds hysterics and paranoids significantly.
[b] Not significantly different from the paranoids.

Hysterics

The hysteric group consisted of twenty women who had been diagnosed "conversion hysteria" at the Illinois Neuropsychiatric Institute. Nineteen of these women were outpatients and one was an inpatient. The age range was from 18 to 50, with the mode in the late thirties. A wide variety of occupations was represented in the group; but the "housewife" category tended to predominate. The educational level of the group was significantly inferior to that of the normals; and a year or two of high school was the most frequently indicated. Most of the group was told that the testing procedures represented part of the routine psychological testing program for patients. However, in some instances due to special circumstances the testing was represented as an "experimental project with particular interest in your type of case." The general level of cooperation in the group was good. In the few instances where subjects showed themselves to be unusually uncooperative testing was discontinued and the results were not used. For about half of the subjects it was necessary to utilize two separate sessions before the test series was completed.

Paranoids

The paranoid group consisted of twenty women psychiatrically diagnosed paranoid schizophrenics. Nine of these subjects were obtained at Elgin State Hospital, five at Chicago State Hospital, and six from the Illinois Neuropsychiatric Institute (three inpatient and three outpatient). The age range was from 21 to 55, with a mode in the late thirties. The educational and occupational representation was similar to that of the hysteric group. The general level of cooperation was significantly poorer than was the case with the other groups. Of course, all subjects used were sufficiently cooperative so that they were willing to undergo a number of hours of testing. Most of the group required at least two sessions and several as many as three sessions before testing could be completed.

CHAPTER III

GENERAL PROCEDURE AND MEASUREMENT TECHNIQUES

A. Procedure

The actual procedure used in the examination of each subject was somewhat as follows:

1. Each individual subject was first given the verbal section of the Wechsler-Bellevue to determine if the intelligence level was within the average range.
2. A Rorschach record was then taken.
3. After the subject had had a brief rest period, the battery of rigidity measures was administered (in the same order as they are described below).
4. Finally, the subject was asked to fill out the Guilford STDCR.

The total time of examination ranged from three hours to seven hours—with a mode of about three and one-half hours. In all instances where subjects seemed fatigued extra rest periods were introduced. The longer examinations were always spaced over a two or three day period.

B. Measurement Techniques

The tasks finally included in the rigidity battery were selected to represent a scatter of situations ranging from those in which the subject deals only with very simple stimuli and a limited number of alternatives, to those in which the subject deals with very complex stimuli and multiple alternatives. Furthermore, they were selected to include tasks varying widely in the degree to which they might "involve" the subject in an emotional sense. Thus, there are two parameters running through the tasks: A *complexity parameter* and an *emotional involvement parameter.* It is important to note that a large number of *overlapping* tasks are employed so that possible failure of any single task to function as a rigidity measure in any individual case will be compensated for by the presence of other overlapping tasks. It should be added that a preliminary test of nineteen tasks was undertaken with ten subjects. Those were retained in the final battery that represented the easiest to administer and seemed to provide the best measure of rigidity.

Each task in the rigidity battery will now be described.

1. *Vigotsky Blocks*

The Vigotsky blocks consist of 22 wooden objects of various forms, colors, and shapes. (16, pp. 462-484). Each subject is instructed: "Show me all of the different ways these blocks can sensibly be divided into groups. Show me one way; then another—until you've shown me all of the ways you can think of." Scoring is based on the number of such groupings he is able to make. The specific details on scoring may be found in Appendix I.

2. *Colors*

Each subject is asked to express his liking for each of eighteen colored ribbons of varied hues and patterns and about four inches in length. As each ribbon is presented separately to the subject, he is asked: "Do you find the color of this ribbon to be pleasing or not pleasing? Do you like it or not like it?" The number of ribbons reacted to favorably constitutes the final score. It is here assumed that limitation in number of positive or favorable reactions is a measure of at least one aspect of rigidity.

3. *Trait Judgment*

The subject is presented successively with several series of pictures of persons—each series consisting of ten pictures. In the first series the subject is told: "Select all those pictured individuals that appear to be friendly." In the second he is instructed to select all those that appear to be ambitious; in the third all those that appear to have a sense of humor; in the fourth all those that appear to be happy; and in the fifth all those that appear to be intelligent. It is assumed that although pictures actually afford some valid criteria for making the above judgments, that the restrictiveness and criticalness of judgment for each series is to a significant degree a function of rigidity trends. The number of pictures accepted as belonging to each trait category constitutes the final score.

4. *Comparison of Self with Pictured Persons*

Series 1.—The subject is shown successively a series of thirty pictures of different women. The subject is then told: "As I show you each picture tell me whether you think the person shown in the picture is a person like yourself—whether she has the same kind of personality as yourself." Additional instructions are given about disregarding any similarities in physical appearance between the pictured person and the self and concentrating on a general impression of the personality of the person represented. The number of pictures that the subject is able to accept or perceive like himself constitutes the total score.

Series 2.—The subject is shown successively fourteen distinctly unpleasant pictures of women taken from the Szondi (24) series and is asked to decide in each case whether the person shown is:

(*a*) "Quite a bit" like the self;
(*b*) "A little bit" like the self;
(*c*) "Not at all" like the self; or
(*d*) Whether it is impossible to reach a decision about the picture in question.

These instructions were given orally; and repeated as many times as necessary until it was clear that the subject comprehended them.

Scoring of the judgments is based on a weighting system that credits +3 for answer "*a*"; +2 for answer "*b*"; +1 for answer "*d*"; and 0 for answer "*c*." It is assumed that the greater the total score, the less the rigidity that is manifested.

5. *Thematic Apperception Test*

a. The subject is shown successively each of ten TAT cards. He is instructed; "Tell me all of the different things this picture might sensibly represent. What are the different things that might be happening in the picture?" Particular care is taken to motivate the subject, to make him feel that he should give as many interpretations as possible. The total score is based on the number of interpretations offered —but qualified in terms of plot quality and also the degree of *difference* of those interpretations given for the *same* TAT card. Details of the scoring system may be found in Appendix II.

b. The subject is shown the same ten TAT cards. In each instance he is asked to select from a *printed list of alternatives* (six per TAT card) all of the various kinds of situations that the card might sensibly represent. The total score is based on the number of alternatives selected by the subject.

6. *Annoyances*

a. Each subject is told: "I am going to read a list of things to you that people sometimes find annoying. As I read each thing to you, tell me whether you find it annoying." A list of twenty-eight annoyances is read to the subject one at a time. The supposition underlying this procedure is that the individual who is least open to grasping alternative ways of interpreting or dealing with situations will tend to react to an unusually large number of aspects of persons and things in an unaccepting (or annoyed) fashion. Clearly, rigidity is here not so directly definable in the form of a simple "number of selected alternatives." Rather, there are more intervening assumptions—the principal of these being that personality rigidity will reflect itself to at least some degree in limitations of outlook and a greater tendency to be rejecting of that which is "different" or new. The total score is the sum of indicated annoyances, and the larger the score the greater the degree of rigidity.

b. Each subject is also asked to select from the same list all those which he thinks "most people" find annoying. When the task is performed with this frame of reference, the total score is based on the number of annoyances ascribed to "most people." The same logic lies behind this procedure as indicated for "self annoyances." However, interest is focused on the rigidity shown when a new, forced frame of reference is involved.

7. *Range of Interests*

a. Each subject is instructed to select from a list of thirty-seven different activities (e.g., swimming, dancing, etc.) which are read to him one at a time all those he feels he enjoys doing *very much*. He is told: "I am going to read a list of activities to you. Tell me which of them you enjoy doing *very much*." The same rationale lies behind this procedure as for the use of annoyances as a measure of rigidity. The total score consists of the number of indicated interests. The greater the number of interests the smaller is the degree of rigidity.

2. Each subject is also asked to select from

the same list all those he feels "most people" enjoy doing *very much*. When the task is performed with this frame of reference, the total score is equal to the number of interests ascribed to "most people."

8. *Rosenzweig Picture Frustration Test* (19)

This test contains a series of cartoon situations in which one person says or does something to frustrate another person. The subject usually writes in a designated box what he thinks the frustrated person would say in response to the frustration. Fourteen of these cartoon situations are used. The subject in each instance selects from a *list of five alternative printed responses* those that he thinks he himself might make under the circumstances. The instructions are: "In each case tell me all of the things in the list that you think you might say if you were in that situation and had to answer the other person." In each printed list the responses are derived from norm lists published by Rosenzweig.

It is clear from Rosenzweig's data that a wide range of responses is possible in each cartoon situation, and so to a large extent restriction as to choice of likely responses is assumed to be a rigidity function. The final score is equal to the number of alternatives selected by the subject.

9. *Blots*

The subject is shown a series of twenty-one ink blots taken from the Terman M-F (25). As he views each blot he is asked to select from a printed sheet containing a listing of possible things the blot might represent (five alternatives per blot) all of the things that he thinks the blot could sensibly be said to represent. The instructions are: "In each case pick out from the list that goes with the ink blot all of those things that you think it could sensibly represent or be." Since the alternatives for each blot are all rather logical possibilities, the number of choices accepted by the subject constitutes the final score.

10. *Hand Steadiness Aspiration Task*

The apparatus used here is a modification of that described by Dunlap (9). It consists of a series of 11 equally spaced holes of decreasing diameters drilled near the periphery of a circular brass plate which can be rotated. The brass plate is mounted on a stand and is at about "chest level" for a subject in a sitting position. A stylus about 15 inches in length is fixed in front of the plate so that its tip is inserted in the third smallest hole. This stylus is on a pivot which allows it to move freely in any direction. The subject holds the stylus with his favored hand in any way he finds "most comfortable." The apparatus appears to be connected to a counter—but actually is not. Consequently, scores can be reported to a subject arbitrarily as the experimenter desires. The instructions to the subject are: "This is a test to see how steady you can hold your hand and also to find out how well you can estimate your own ability to do things. With your eyes closed try to hold this rod in the middle of the hole as steady as you can. At the end of each trial I will tell you how many times you actually touched the sides of the hole. Then, you will try to estimate in advance the number of times you will touch it on the next trial." He is then given a *practice* trial for thirty seconds with his eyes closed.

After the practice trial, ten *regular* trials are run. Before each trial the subject estimates what he expects to achieve; and at the end of the trial an arbitrary score, and yet one reasonable in terms of the possible achievement range, is reported to him. Following the first trial and for three further successive trials the subject is permitted to improve his performance progressively in terms of a predetermined pattern. Failure to raise or lower aspiration estimates is taken as the scoring basis. The reasoning underlying the use of this task is that on each trial the subject has the choice of a theoretically wide range of estimates clustering somewhere near the "real score" reported by the experimenter. Consequently, any resistance he shows to the acceptance of the full range of possibilities and any predilection he shows for "sticking to" his own *previous* estimates serve as an index of his inability to utilize a range of alternatives in his reactions to situations (especially those of a stress character). Three scores are derived from this task: (1) One indicating the amount of resistance to change when progressive improvement is reported to the subject; (2) One indicating the amount of resistance to changes when progressive decline in skill is reported; and (3) One indicating the degree of difference between rigidity for the first phase as contrasted with the second phase. See Appendix III for details of the scoring procedure.

CHAPTER IV

ANALYSIS OF DATA

A. Validity Criteria

1. *Rorschach*

In using the Rorschach test as a validity criterion against which to compare results obtained from the rigidity battery, it was decided to work out three separate measures:

a. A measure of personality rigidity.

b. A measure of degree of total maladjustment.

c. A measure of degree of pathological looseness or disorganization.

There are different ways in which one can approach the problem of evaluating Rorschach material. When using the Rorschach clinically, one usually attempts to grasp the various numerical scores obtained, in terms of total broad patterns or sub-patterns. Actually, the Rorschach literature contains abundant evidence that this is the most fruitful mode of analysis. However, since broad patterns are extremely difficult to express quantitatively, most controlled studies involving the Rorschach test have tended to rely on correlations of *single* scores (out of context of the total pattern) with other variables. Such analysis of isolated scores has generally been very disappointing and for this reason an attempt has been made here to define the Rorschach measures so that they retain some of the quality of *pattern* evaluation. This has been done by working out a scheme whereby individual scores are given different weights in terms of their quantitative relation to other scores.

The above procedure will become clearer as the specific scoring systems for evaluating rigidity maladjustment and pathological looseness are described in detail. The deficiencies of such scoring systems must of course be kept in mind constantly. At most, they are only rough techniques of evaluation. They only touch on the complexity of analysis that goes into a careful *qualitative* consideration of a Rorschach record. Many complicated questions regarding relations of Rorschach factors would have to be worked out carefully in any project directed at Rorschach quantification per se; but these questions have necessarily received only a moderate degree of attention here. However, as will later be indicated, all of the scoring systems seem to be sufficiently differentiating to have possibilities as validity criteria.

a. The Rorschach measure of *personality rigidity* is based on a variety of Rorschach signs which clinically have been found to characterize persons who are habitually constricted and who find it necessary to deal with the environment with an unusual amount of guarded caution. Quite literally, these signs are derived from indicators of the degree to which individuals limit the range and character of their responses to the Rorschach ink blots. Thus, limitation of number of responses; limitation in the use of color in perceptions; restriction of percepts to one class of objects (e.g., animals); and very delayed reaction times are illustrative of what is considered rigid or restricted behavior in the Rorschach test. To construct a measure which would objectively convey the amount of Rorschach rigidity shown, a variety of the above described Rorschach signs were defined in terms of various scores (4) and given a series of specific weights. The particular rigidity signs selected and the weights assigned them were determined in an arbitrary manner based on general clinical experience. It is not possible to demonstrate that the signs or weights selected are the truest measure of personality rigidity in the Rorschach sense. They were chosen to express what was qualitatively suggested in the Rorschach data obtained. It is likely that better weighting systems could be worked out; the one presented here is merely the best that could be developed under the conditions of this project. The value of this procedure is that it gives an independent measure of rigidity based on an instrument which has proved clinically useful. Thus, some frame of reference for evaluating the *rigidity battery* results is provided. The details of the scoring system may be found in Appendix IV.

The differences in Rorschach rigidity among the clinical groups are shown in Table 2. It may be noted that the normal group is significantly less rigid than either the hysteric or paranoid groups.

TABLE 2

CONSTANTS OF THE DISTRIBUTION OF RORSCHACH RIGIDITY SCORES IN THE THREE EXPERIMENTAL GROUPS, AND DIFFERENCES IN RORSCHACH RIGIDITY AMONG THESE GROUPS

	Normal	Hysteric	Paranoid
Mean	24.9	44.3	44.1
σ	11.8	17.2	13.6
σ_m	2.7	3.9	3.1
	Normals vs Hysterics	Normals vs Paranoids	Hysterics vs Paranoids
Mean Difference	19.4	19.2	.2
σ_D	4.7	4.1	5.0
C.R.	4.1	4.7	.1

The fact that no significant difference exists between the gross scores of the paranoids and conversion hysterics is consistent with results recently obtained by Angyal (3).

b. The Rorschach maladjustment score is also derived in terms of pattern criteria. A range of Rorschach signs which are clinically recognized as indicating various degrees of personal maladjustment were selected. Each sign was given a weight in terms of its relation to other factors and its assumed importance for diagnosing maladjustment. Details of the scoring may be found in Appendix V. The manner in which the Rorschach maladjustment score differentiates the clinical groups is shown in Table 3.

The normal group is very well differentiated from both the hysterics and paranoids. However, it is clear that the hysterics and paranoids are less well differentiated, although still significantly in a statistical sense. Aside from defects in the scoring system used, some of the overlap may be attributed to the following factors:

1. Several of the hysterics are extremely maladjusted—perhaps bordering on incipient psychosis.
2. A number of the paranoids had been hospitalized for acute episodes which had occurred some time previous to the time of study and so might well have attained various stages of recovery.
3. It is generally recognized that paranoid disorders are difficult to detect on the Rorschach test—even when careful qualitative analysis is utilized.

TABLE 3

CONSTANTS OF THE DISTRIBUTION OF RORSCHACH MALADJUSTMENT SCORES IN THE THREE EXPERIMENTAL GROUPS, AND DIFFERENCES IN RORSCHACH MALADJUSTMENT AMONG THESE GROUPS

	Normal	Hysteric	Paranoid
Mean	36.9	59.7	85.5
σ	11.0	19.1	32.4
σ_m	2.5	4.3	7.4
	Normals vs Hysterics	Normals vs Paranoids	Hysterics vs Paranoids
Mean Difference	22.8	48.6	25.8
σ_D	2.8	7.7	7.5
C.R.	8.1	6.3	3.4

In general, then, it would appear that the Rorschach maladjustment score is able to detect personality maladjustment with at least moderate efficiency and that one may feasibly use it as a "maladjustment criterion" for interpreting the results of the rigidity battery.

c. The *"looseness"* or *pathology* score is a separate compilation of all penalty weights given in the *maladjustment scoring* which are judged to indicate looseness or disorganization in the personality structure. It seemed desirable to have a separate index which would point up more serious pathology, such as peculiarities in thinking logic or unusual perseverative interest in sexual topics. This score is used only for a better understanding of the maladjustment score. Details relative to the penalty weights which go into this score may be found in Appendix VI.

An attempt was made to derive a number of other Rorschach scores (e.g., *degree of general hostility* and *reaction to stress*). Because of their limited range of variation and their general lack of promise they were not given further consideration.

2. *Guilford's STDCR*

This inventory was not used with the intention of actually measuring specific real personality factors such as S, T, D, C, R that the device purports to analyze and then comparing them with rigidity battery results. The main purpose in utilizing the STDCR was to have a uniform means of analyzing a subject's way of describing himself. That is, all of the questions in the STDCR inquire about some behavioral or subjective aspect of the subject; and the way he answers these questions constitutes a form of self description. In answering the various questions he is implicitly saying such things as: "I feel happy (or sad)." "I have many (or few) friends." "I am calm (or excitable," etc. It is recognized, of course, that in some instances a subject may have so much insight into himself that his self descriptions are actually valid indicators of his real personality. However, the fact of these descriptions being *self descriptions* remains.

With this point in mind, the highness or lowness of the total of the five factor scores is the chief STDCR measures used. A high score means that the subject describes himself in favorable or socially "good" terms. It means that he refers to himself as happy, well adjusted, satisfied, etc. A low score indicates that the subject describes himself in unfavorable "poor adjustment" terms. No assumptions regarding the motivation for these different kinds of self descriptions are made. The STDCR scores will be treated simply as empirical indicators of favorable or unfavorable self description.

It was noted that an additional point to be taken into account when interpreting the STDCR total as a mode of "self description" was the amount of variation in the STDCR scores. That is, a high total accompanied by moderately high variation seemed to have a different significance than a high total accompanied by low variation. For this reason a *variation score* for each series of STDCR scores was worked out. If the five STDCR scores are considered to be ranked in order from highest to lowest ("favorable" self description to "unfavorable" self description), the variation score is equal to the difference of the fifth score from the fourth score *plus* the difference of the third score from the second score *plus* twice the difference of the fifth score from the first score.

This specific way of computing variation is an attempt to maximize those aspects of the variation which seemed most important (in terms of qualitative inspection) for understanding the highness or lowness of the STDCR totals. Further discussion of the variation score will occur at a later point.

That the STDCR total in this study is largely not a measure of real personality factors but is at least to some degree only a subjective kind of self description is implied by the results in Table 4. Thus, as can be seen from Table 4, the paranoid schizophrenics do not differ significantly from the normal subjects. This is true also when the factor scores are considered individually. Furthermore, although the paranoids are clinically and in terms of test criteria *more disturbed* than the conversion hysterics, their STDCR scores would misleadingly imply that their personality traits and reactions are *more appropriate* and adjustive than those of the hysterics. Further evidence in a similar vein will be presented in more precise terms at a later point.

TABLE 4

CONSTANTS OF THE DISTRIBUTION OF *STDCR* TOTAL SCORES IN THE THREE EXPERIMENTAL GROUPS, AND DIFFERENCES IN *STDCR* TOTAL SCORES AMONG THESE GROUPS

Constants	Normal	Hysteric	Paranoid
Mean*	32.1	20.9	28.2
σ	5.1	7.2	8.1
σ_m	1.2	1.6	1.8
Constants	Normals vs Hysterics	Normals vs Paranoids	Hysterics vs Paranoids
Mean Difference	11.2	3.9	7.3
σ_D	2.0	2.2	2.4
C.R.	5.6	1.8	3.0

* The higher the mean, the "better" the self description implied.

3. *Clinical Diagnosis*

It should be briefly noted that when this project was first undertaken conversion hysterics and paranoids were included on the widely held assumption that the hysterics would represent at one extreme the non-rigid type of personality and the paranoids at the other extreme a very rigid type of personality. It was thus that a clinical validity criterion for judging the rigidity battery would have been provided. However, the obtained results (verified by another study, 3, pp. 133-135) indicate that conversion hysterics and paranoids lie at the same extreme of rigidity. Consequently, validation comparisons are restricted mainly to contrasts between the normals on one hand and the conversion hysterics and paranoids on the other hand.

B. PRELIMINARY ANALYSIS OF INDIVIDUAL TASKS IN THE RIGIDITY BATTERY

This first phase of analysis is intended to describe very briefly some of the characteristics of the *individual* tasks in the rigidity battery: to make clear in a general fashion how the individual tasks are related to a variety of validity criteria and to point out the limitations of their independent use.

It is quite clear from the data obtained that *no single rigidity measure in the total battery is very useful in terms of predictive efficiency.* Most of the correlations and comparisons with validity criteria given below are definitely low—although there is a certain trend and consistency in relationship that appears throughout the results. Thus, (as shown in Table 5), only two tasks (Vigotsky sorting and spontaneous TAT interpretation) *significantly* differentiate the normals as being less rigid than both the hysterics and paranoids.

TABLE 5

CONSTANTS OF THE DISTRIBUTIONS OF INDIVIDUAL RIGIDITY TASK SCORES IN THE THREE EXPERIMENTAL GROUPS, AND DIFFERENCES IN INDIVIDUAL RIGIDITY TASK SCORES AMONG THE GROUPS

Rigidity Task	Mean of Normal	Mean of Hysteric	Mean of Paranoid	σ of Normal	σ of Hysteric	σ of Paranoid	N[a] vs. H			N vs. P			H vs. P		
							Diff.	C.R.[b]	Diff. Favors	Diff.	C.R.	Diff. Favors	Diff.	C.R.	Diff. Favors
Vigotsky	16.1	10.3	11.3	4.9	3.9	4.7	5.8	4.1	H	4.8	3.2	N	1.0	.8	P
Colors	11.5	12.2	13.3	2.4	2.8	3.8	.7	.9	H	1.8	1.8	P	1.1	1.1	P
Traits	20.9	19.1	21.4	5.5	5.9	7.8	1.8	1.0	N	.5	.2	P	2.3	1.0	P
Self (ordinary series)	11.1	7.7	5.4	5.1	4.9	4.7	3.4	2.1	N	5.7	3.1	N	2.3	1.5	H
Self(unusual series)	8.3	5.4	4.8	5.8	5.6	5.2	2.9	1.6	N	3.5	1.9	N	.6	.4	H
Independent TAT	28.5	22.3	22.6	5.2	5.2	6.8	6.2	4.2	N	5.9	3.1	N	.3	.2	H
TAT choice	27.9	24.2	22.3	7.5	8.9	10.6	3.7	1.4	N	5.6	1.9	N	1.9	.6	H
Annoyance (S)	13.9	16.4	12.3	6.0	3.9	6.9	2.5	1.5	H	1.6	.8	N	4.1	2.0	H
Annoyance (O)	18.4	19.7	17.4	7.4	4.0	6.6	1.3	.7	H	1.0	.6	N	2.3	2.3	H
Interests (S)	17.4	14.2	16.8	5.9	5.4	6.3	3.2	1.8	N	.6	.3	N	2.6	1.4	P
Interests (O)	20.1	22.3	21.0	6.2	5.1	7.4	2.2	1.2	H	.9	.4	P	1.3	.7	H
P−F	30.9	28.9	26.9	6.7	11.0	9.5	2.0	.7	N	4.0	1.5	N	2.0	.6	H
Blots	48.6	39.5	35.1	16.2	14.4	11.2	9.1	1.9	N	13.5	3.0	N	4.4	1.1	H
Aspiration (S)	+6.3	+5.0	+4.2	4.2	4.2	3.9	1.3	.9	N	2.1	1.6	N	.8	.6	H
Aspiration (F)	+2.2	−1.0	−.6	4.6	6.0	6.6	3.2	1.9	N	2.8	1.6	N	.4	.2	H

[a] N=Normal; H=Hysteric; P=Paranoid.

[b] C.R.=Critical Ratio. A critical ratio of 3 has been adopted as the level of significant difference. This high standard of significance has been chosen because it is felt that the rather new and sometimes vaguely defined variables here encountered should be approached with special caution in order to prevent too easy speculation about them.

But at the same time in twenty-three out of thirty instances in which normal means are compared with either hysteric or paranoid means the normals are indicated to be less rigid; and similarly of the thirty critical ratios derived from these comparisons thirteen are 1.8 or higher. Although eleven out of fifteen mean comparisons favor the hysterics as less rigid than the paranoids, these differences are so small and unreliable that no real conclusions on the point can be drawn.

When the correlations of single task rigidity scores with Rorschach *rigidity scores* are computed (see Table 6) results are similar to those obtained from the comparison of the means. That is, very few significant correlations appear, although there is a certain trend of relationship throughout. Thus, no single score correlates well with Rorschach rigidity in every group; and only two tasks (self interests and spontaneous TAT interpretations) correlate .40 or higher for the three groups combined. However, five out of eleven correlations for the *total* group are .30 or higher. Sixteen of the *individual* group correlations are .20 or higher and seven of these correlations are .40 or higher. Here too, the fact of low relationship is outstanding.

A general comparison of single task scores with *Rorschach maladjustment* scores further emphasizes the limitations of these measures of simple tasks when used independently. As can be seen, no given task (Table 7) correlates significantly with Rorschach maladjustment in all of the diagnostic groups. One is struck by the peculiar range of coefficients. Rigidity on some tasks correlates positively with degree of maladjustment and in other instances rigidity correlates negatively with maladjustment. Indeed, when one observes, for example: (*a*) that flexibility in color preference has a —.66 relation to Rorschach maladjustment in the paranoid group but a +.16 relationship in the normal group; (*b*) or that low rigidity in terms of self interests has a —.55 relation to Rorschach maladjustment in the hysteric group, but a +.25 relationship in the paranoid group —it may be assumed that the relationship of rigidity to maladjustment is probably a complex one that will not reveal itself in these single isolated measures. Some of the factors underlying the range of

TABLE 6

CORRELATIONS* OF INDIVIDUAL RIGIDITY SCORES AND RORSCHACH RIGIDITY SCORES

Variables	Normals	Hysterics	Paranoids	Total group
	Rho	*Rho*	*Rho*	*Rho*
Colors	—.13	—.18	—.06	—.19
Traits Judgment	—.14	—.25	+.28	+.16
Pictures Like Self	+.06	+.18	—.10	+.17
Self Annoyances	—.20	+.37	—.05	+.12
Other Annoyances	—.19	—.25	—.01	—.09
Self Interests	+.41	+.56	+.05	+.42
Other Interests	+.21	+.09	—.51	—.11
Independent TAT	—.02	+.76	—.08	+.49
P—F	+.17	+.58	—.20	+.34
Blots	+.16	+.58	+.26	+.34
Aspiration (F)	+.46	+.22	—.02	+.30

---

* The correlations in this table and in all tables to follow are *Rho* correlations.

TABLE 7

Correlations of Individual Rigidity Tasks with Rorschach Maladjustment

Variables	Normals	Hysterics	Paranoids	Total Group
	Rho	*Rho*	*Rho*	*Rho*
Color	+.16	−.20	−.66	+.16
Trait Judgment	+.51	+.02	+.29	+.13
Pictures Like Self (usual series)	+.60	+.13	+.38	−.10
Annoyances (S)	+.39	−.54	+.03	+.04
Annoyances (O)	+.52	−.07	+.12	+.03
Interests (S)	−.13	−.55	+.25	−.19
Interests (O)	−.17	−.12	−.12	−.02
Independent TAT	−.21	+.12	+.42	−.28
Choice TAT	−.17	−.20	+.38	−.19
P−F	−.34	−.34	+.20	−.30
Blots	−.24	−.08	+.34	−.23
Aspiration (F)	−.49	+.02	−.24	−.32

correlations here obtained will become clearer at a later point when more complex analyses of the data are described.

Altogether, then, one may conclude that there is little to be gained by attempting to measure or understand rigidity and its related phenomena purely in terms of each separate task comprising the rigidity battery.

C. Analysis of Total Rigidity Score

The *total rigidity score* is equal to the sum of an individual's ranks for sixteen subtest scores. This general measure proves to be a more satisfactory one than any of the single task measures. It has also proved itself more effective than various combinations of individual tasks that were attempted. Its differentiations and correlations relative to validity criteria are at a more significant level. Before preceding to an analysis of these relations with validity criteria it will be well to deal with the question of whether the single task scores show some kind of consistent relationship to the total score and to what degree they overlap in their contribution to the total score.

The single scores show a low to moderate relationship to the total score. Thus, for the *normal* group only six of the twelve correlations are .30 or higher and only three are .40 or higher. Likewise, for the *hysteric* group seven of the twelve correlations are .30 or higher and four .40 or higher. In the *paranoid* group one finds only four correlations that are .30 or higher and only one correlation equals or exceeds .40. There are distinct

TABLE 8[a]

Correlations of Single Rigidity Scores with Total Rigidity[b]

Variables	Normal	Hysteric	Paranoid
	Rho	*Rho*	*Rho*
Color	+.27	−.26	−.02
Traits	+.03	+.24	+.34
Self (usual)	+.37	+.38	+.36
Annoy. (S)	+.21	+.13	−.31
Interest (S)	+.32	+.56	+.05
Annoy. (O)	+.19	−.22	−.05
Interest (O)	+.12	+.22	+.04
TAT (ind.)	+.53	+.48	+.25
TAT (Choice)	+.52	+.67	+.28
P−F	+.46	+.38	+.17
Blots	+.33	+.52	+.56
Aspiration (F)	+.23	+.30	+.17

[a] Four of the sixteen individual tasks are not included in this analysis because their restricted scatter made correlations rather meaningless.

[b] In each case where a total score was correlated with a specific individual score, the individual score had previously been subtracted from that total score.

differences in the degree to which the single tasks seem to measure whatever there is that is common to the total score. Thus, color preferences, self annoyances, annoyances of others, and interests of others tend with but a few exceptions to give very low coefficients. On the other hand, such tasks as independent TAT interpretation, choice TAT interpretation, and blot identification tend to range at a distinctly more significant level. However, it is clear that in general the single tasks do not measure any simple common uni-dimensional factor in any real sense. The significance of this fact will be clarified and expanded at a later point.

To determine how much overlap or duplication there was among the various individual tasks, scatter-diagrams were plotted in all those instances where inspection of the data might lead one to anticipate significant relationships. Correlation coefficients were then determined for those scatters that suggested significant overlap. Although there are numerous instances in which specific tasks give moderately high correlations within the setting of a given diagnostic group and although there are several cases in which moderately high correlations obtain when the three diagnostic groups are combined, there is no instance in which the overlap between two tasks persistently appears in *each* of the diagnostic groups. Such data tend to imply that none of the tasks duplicate each other to such an extent that it would be considered advisable to let one represent another. A more specific indication of the manner in which the tasks do overlap is shown by Table 9.

Most of these correlations are below .50; and only three of them are .60 or higher. (It will become clear at a later point that moderate to high overlap between certain task scores is advantageous in the measurement of rigidity tendencies.)

Returning to the question of what kind of relationships obtain between total rigidity scores and criterion scores, we find that when groups are compared (see Table 10) in terms of their mean total scores the normals are significantly differentiated as being less rigid than either the hysterics or the paranoids. Once again there is no real difference between the hysterics and paranoids. Indeed, it is interesting to note how similar the means of the two groups are. Among the *single* rigidity tasks, it is true that

TABLE 9

MOST SIGNIFICANT CORRELATIONS AMONG THE INDIVIDUAL RIGIDITY TASKS

Variables	Normals		Hysterics		Paranoid	
	Rho	S.E.	*Rho*	S.E.	*Rho*	S.E.
Interest S+O	+.56	.16	N.S.*		N.S.	
Annoyances S+O	+.89	.07	N.S.		+.64	.14
Independent TAT+Choice TAT	+.72	.10	N.S.		N.S.	
Traits+Self (Ordinary Series)	+.43	.19	N.S.		N.S.	
Blots+P−F	+.48	.18	N.S.		N.S.	
TAT Choice+Interest (S)	N.S.		+.51	.18	N.S.	
Blots+Interests (S)	N.S.		+.49	.19	N.S.	
Blots+Traits	N.S.		N.S.		+.48	.18
Blots+Aspiration (S)	N.S.		N.S.		+.49	.18

* N.S. = Not significant.

TABLE 10

Constants of the Distributions of Total Rigidity Scores in the Three Groups, and Differences in Total Scores Among These Groups

Constants	N*	H	P	N+H	N+P	H+P
M	418.0	534.0	517.0			
σ	91.5	110.0	93.2			
σ_m	20.8	25.2	21.1			
D				116.0	99.0	17.0
σ_D				33.0	29.3	33.0
C.R.				3.5	3.4	.5

* N = Normal; H = Hysteric; P = Paranoid.

Vigotsky sorting and spontaneous TAT interpretation differentiate the normals from the hysterics and paranoids at a more significant level than the total score; but the superiority of these tasks disappears when another criterion such as Rorschach rigidity is brought into the picture.

The relation of the total rigidity score to *Rorschach rigidity* tends to be moderately high. It ranges from .10 in the paranoid group to .73 in the hysteric group. For the three groups combined the coefficient is .52. These differences are probably at least in part due to the narrower range of scores in the normal and paranoid groups as contrasted with the hysteric group. But there are more important reasons for these differences which will be discussed at a later point. Altogether, the preceding data imply that total rigidity scores measure *"general"* differences in rigidity trends with only low to moderate success.

TABLE 11

Correlations of Total Rigidity Score with Rorschach Rigidity

Normal	Hysteric	Paranoid	Total
+.31 S.E. (.21)	+.73 S.E. (.11)	+.10 S.E. (.24)	+.52 S.E. (.10)

Although only those persons who fell within the IQ range of 95 to 110 were used as subjects, still IQ differences in the groups exist. One might ask whether the observed *general* rigidity trend differences are not basically due to variations in intelligence.

The correlations shown in tables 12 and 13 indicate that intellectual differences do contribute in a limited degree to the total rigidity score. However, as can be seen from the correlations of intelligence with Rorschach rigidity, the contribution which intelligence makes to the total score seems in general to be different from that which the Rorschach rigidity scores and the total scores have in common.

TABLE 12

Correlations of Intelligence with Total Rigidity Score

Normal	Hysteric	Paranoid	Total
+.02 S.E. (.24)	−.29 S.E. (.22)	−.52 S.E. (.18)	−.51 S.E. (.10)

And if one assumes that Rorschach rigidity is basically some kind of index

TABLE 13

Correlations of Intelligence with Rorschach Rigidity

Normal	Hysteric	Paranoid	Total
+.01 S.E. (.24)	−.27 S.E. (.22)	−.04 S.E. (.24)	−.28 S.E. (.12)

of personality rigidity, it would appear that the total rigidity score does get at a personality rigidity variable which is not measured by intelligence. The fact that intelligence does in some of the groups have a significant influence on *general rigidity trends* is not surprising when one considers the interdependence of intellectual and emotional factors; and remembers that the intellectual assets of an individual may aid him in solving very personal emotional problems or benefiting from psychiatric therapy. Generally, the total rigidity score refers to rigidity trends that are fairly independent of intelligence; but it is likely that, if one wanted to compare individuals of widely differing intelligence, separate norms for the different intellectual levels would be necessary.

Some effort has been directed to demonstrating that the battery of rigidity tasks measures rather consistently some common variable that may be called "personality rigidity," and the indications are that it does measure such a variable to a certain degree. However, as one reviews the data and notes such facts as unusual shifts in correlation from group to group; inconsistent relationships; and the wide scatter of *single task scores* within many subjects' total scores, two possibilities come to mind:

a. One may hypothesize that the present technique for measuring *personality rigidity* has only a moderate efficiency and that there are many situations in which it is very ineffective as a measuring device.

b. Or, one may postulate that to conceive *personality rigidity* as a measurable generalized characteristic of personality structure is fictional over-simplification. That is, one may assume that there are general *tendencies* toward rigidity or looseness in most personality structures; but that these are only tendencies and that possibly there are different levels of rigidity and perhaps even contradictory currents of rigidity in the same individual. Certainly the first hypothesis accounts for some of the discrepancies in results, but it will be shown in the next section that the second hypothesis seems most applicable.

D. Analysis in Terms of Levels of Rigidity

If one prepares profiles of the various task scores for each individual, one is struck by the fact of *variation* rather than of uniformity. It is the rule to find that most subjects show a wide range of differences in their separate scores. The individual who is uniformly high or low is an exception. This is not at all strange when one considers the range of difficulty and emotional stress represented among the tasks in the rigidity battery. Some tasks (e.g., color preference or trait judgment) simply call for an expression of opinion and permit the subject to reach a decision essentially in terms of his own frames of reference without real concern as to the rightness or wrongness of his decision. Others (e.g., independent TAT interpretation and aspiration estimation during failure) place the subject in a position where he has to decide about things that have some degree of emotional importance to him and where he feels that his decisions are more or less subject to external criticism. Is it not assuming an unusually simple formulation of personality to expect that rigidity shown for the first type of task will have any positive relation to rigidity shown in the second type of task? With this question in mind it was decided to work out a rough arbitrary classification of the rigidity tasks into a series of levels and to determine what kinds of relations obtained among these levels over the range of subjects. Finally, the task was seen as one of establishing the significance of these relationships of levels by using the validation criteria available.

The rigidity tasks may be roughly classified into three levels—as defined by the criteria that follows:

Level I	*Level II*	*Level III*
The task represents a demand on the subject which he finds difficult to deal with because it arouses emotion or calls for a type of response which is not readily available to him in terms of his background of experience	The task involves a moderate degree of difficulty, but is at the same time subject to considerable "private frame of reference" determination.	The task represents a demand to which the subject can respond *without much difficulty* or in very arbitrary unchallengeable terms dictated by his own private frames of reference. In other words, he may respond to such tasks about "as he pleases."

In terms of these criteria the specific tasks may be considered to group themselves in the following categories:

Level I

1. Aspiration estimates during failure.
2. Independent TAT interpretation.
3. Choice TAT interpretation.
4. Blot identification.

Level II

1. Self interests.
2. Interests of others.
3. Self annoyances.
4. Annoyances of others.
5. Aspiration estimates during success.
6. Vigotsky sortings.
7. Responses to the P-F situations.

Level III

1. Color preferences.
2. Selection from ordinary range of pictures of persons like the self.
3. Selection from unusual range of pictures of persons like self.
4. Judgment of traits from pictures.

The grouping of tasks into levels is assumed to hold true only "on the average." Thus, it is possible that for some subjects a level II task is actually a Level I task. However, the presence of several tasks at each level minimizes error of this kind.

When the single rigidity scores of each subject were viewed in terms of the above scheme, it was evident that there were definite patterns of relationship among them which could be precisely defined. Furthermore, almost immediately after such patterns were first observed it was noted that certain of them were associated with individuals who might be described as well adjusted (in terms of Rorschach and clinical criteria) and those personality structures appeared to be integrated in a balanced fashion; whereas other patterns were associated with persons who were ill and disturbed. The pattern associated with the adjusted integrated person may be pictured somewhat in the fashion shown at the top of the next page.

The integrated personality is marked by lowest degree of rigidity at level I, with moderately increasing rigidity at levels II and III. It is also characterized by a definite tendency for single tasks at the *same level* to indicate a similar degree of rigidity. However, the pattern of the disturbed individual frequently departs radically from the "healthy" pattern. Greatest rigidity may be shown at level

Characteristics of the Rigidity Levels in "Ideal Profile"

Level I	*Level II*	*Level III*
(A) Of the three levels this is *the least rigid.*	(A) Occupies middle point of rigidity.	(A) Highest rigidity.
(B) All rigidity scores at this level cluster at a similar point of rigidity.	(B) All rigidity scores at this level cluster at a similar point.	(B) All rigidity scores cluster at similar point.

I instead of level III; or tasks at the same level may give widely different rigidity scores. In order to test out the above suppositions in precise fashion it was first necessary to construct an "ideal healthy profile," in which the relationships of representative tasks from each level were quantitatively defined. Furthermore, it was necessary to work out a system of weighted penalties to score the deviations of each subjects profile from this "ideal profile." There follows a summary of the scoring scheme that was evolved.

As can be seen in the above summary, scoring weights are assigned so as to penalize all deviations from the "ideal profile" and to give positive credit for conformance to it. In sections A and B of the scoring summary credits are given in terms of the assumption that rigidity for level I tasks should be *less than* rigidity for level II and III tasks. Each level is represented by sample tasks rather than by all the tasks falling at that level. *Rank* differences between the level I and level II and III tasks in the optimum direction determine the positive weighting and rank differences contrary to the optimum direction determine the negative weighting. In section C an evaluation is made of the degree to which level II rigidity is less than level III rigidity, by comparing sample tasks from each of these levels. Finally, in section D a test is made of the extent to which sample tasks from the same rigidity level actually give similar rigidity scores. A fixed positive weight is assigned for each instance in which the subject responds with equivalent rigidity to tasks belonging to the same level. A range of negative weights is used to penalize significant *differences* in rigidity among tasks of the same level.

The final total score is the difference between the sum of the positive weights and the sum of the negative weights.

The scores secured through the above procedure manifest some very interesting relationships to maladjustment, as defined by the Rorschach test. Within the normal group a very clear correlation obtains. The fact of clear correlation is also evident when the three separate groups are treated as one. However, both the hysteric group and the paranoid group give insignificant coefficients for the two variables. When these two groups were analysed to determine why they differed so markedly from the normals it was discovered that in both instances the correlations were thrown off by several cases that had certain important characteristics in common. It was found that these particularly deviant cases were all marked by the following:

1. They were all among the most extremely sick and disorganized of their respective groups —as defined by Rorschach maladjustment scores, Rorschach looseness scores, and clinical symptoms.

2. They were all deviants from the correlation trends of their respective groups in the sense that their "balance" rigidity scores implied unusually little disturbance—whereas their Rorschach scores implied that they were very ill. On the basis of the above, a hypothesis was

TABLE 13A

SUMMARY OF METHOD FOR SCORING RIGIDITY PATTERN DISTURBANCE

Optimum Relationship	Ranks	Positive Credit	Ranks	Negative Credit
A. 1. Rigidity of aspiration (F) *less than* rigidity of color preference 2. Rigidity of aspiration (F) *less than* rigidity of annoyances (S) 3. Rigidity of aspiration (F) *less than* rigidity of trait judgment 4. Rigidity of independent TAT *less than* rigidity of color preference 5. Rigidity of independent TAT *less than* rigidity of annoyances (S) 6. Rigidity of independent TAT *less than* rigidity of trait judgment	If: 1–2 ranks less 3 ranks less 4–5 ranks less 6 or more ranks	 +1 +2 +3 +4	If: 1–2 ranks more 3–4 ranks more 5 or more ranks	 −1 −2 −3
B. 1. Rigidity of aspiration (F) *less than* rigidity of aspiration (S)	If: 1 rank less 2 ranks less 3 ranks less 4 ranks less More than 4 ranks	 +1 +2 +3 +4 +5	If: 1–2 ranks more 3–4 ranks more 5–7 ranks more 8 or more ranks	 −1 −2 −3 −4
C. 1. Rigidity in selecting pictures like self (unusual series) *more than* rigidity relative to self interests *and* Rigidity in responding to P–F situations	Same weights as used in (A) above		Same weights as in (B) above	
D. Following pairs of tasks should be at same rigidity level[a] 1. Aspiration (F) and independent TAT interpretation 2. Aspiration (F) and choice TAT interpretation 3. Aspiration (F) and blot identification 4. Independent TAT and choice TAT interpretation 5. Choice TAT interpretation and blot identification 6. Self interests and interests of others 7. Acceptance of "ordinary" pictures like the self and "unusual" pictures like the self[b]	—	+4	—	For each rank difference beyond two, negative weights are assigned as in (B) above

[a] That is, within two ranks of each other.

[b] Final score is equal to sum of positive and negative credits.

LIBRARY
12.8.70
9155

TABLE 14

CORRELATIONS OF "BALANCE RIGIDITY SCORES" AND RORSCHACH MALADJUSTMENT SCORES

Normal	Hysteric	Paranoid	Total Group
+.84 S.E. (.07)	+.06 S.E. (.24)	+.19 S.E. (.23)	+.42 S.E. (.11)

formulated that the balance rigidity score was not a measure of maladjustment, as such; but rather a measure of *present degree of disturbance* in an individual's rigidity defenses—that is, a measure of how "comfortable" or "disturbed" he was with the particular style of defense or adjustment he was now utilizing. This would perhaps explain the deviant cases—since there is some logic in assuming that individuals who manifest very primitive symptoms or extreme disorganization have ceased to struggle with their difficulties and are no longer subject to the tensions associated with struggle.

To test this viewpoint, it was decided to continue the use of the Rorschach maladjustment score as a measure of present disturbance in adjustment. However, it was first necessary to isolate and eliminate objectively all cases from the hysteric and paranoid groups that might be considered to have "settled" in rather regressive fashion for a specified neurotic or psychotic adjustment, and whose Rorschach maladjustment scores would consequently not represent an adequate index of disturbance in defenses. This was done by eliminating any case in which the rigidity balance score implied that the individual was among the twenty per cent of *least disturbed* individuals in his group, but whose Rorschach maladjustment score or Rorschach looseness score (whichever was applicable) indicated that he was among the twenty per cent of most maladjusted or disorganized in his group. By this procedure three hysterics and two paramoids were eliminated.

It is interesting to note that the three hysterics selected have identical symptoms, viz., they are all subject to generalized tremors or convulsive-like seizures involving the entire body. But it is most surprising of all to find that they are the only hysterics to manifest such behavior as their dominant symptomatology. These symptoms are obviously of a more regressive and primitive character than the headaches, pains, and slight paralyses characteristic of the other hysterics in the group. The two paranoids eliminated were clinically both marked by extreme looseness. Their Rorschach protocols are pervaded by an extreme amount of anal and sexual content (e.g., "bloody vagina," "penis," "hemmoroid pretrusions") which was expressed quickly and without hesitation. One of the paranoids spontaneously remarked that she was "hearing voices" and both were passively compliant during examination—all of which is in definite contrast to the reserved hostile behavior shown by most of the other subjects in the paranoid group. In general, one seems justified in concluding that the cases eliminated in terms of the contrasting extremeness of their "balance" and Rorschach scores show certain clinical uniformities that confirm earlier stated assumptions about the level of their adjustment.

When correlations between rigidity balance scores and Rorschach maladjustment scores are determined on these "clean" groups, the correlation for the hysteric group shifts from +.06 to +.46 and the correlation of the paranoid group shifts from +.19 to +.72. The correlation for the total group shifts from +.42 to +.60. That is, if the rigidity balance score[1] is conceived as a measure of present disturbance in defenses it correlates with suitably interpreted Rorschach maladjustment scores quite highly.

The diagnostic possibilities of the rigidity balance score and its usefulness in understanding the dynamics of individual cases will be discussed at a later point.

Altogether, the data presented in this section may be said to carry the following implications:

---

[1] It may be mentioned here that in no instance do rigidity balance scores show any relationship to intelligence scores.

TABLE 15

SUMMARY OF CORRELATIONS BETWEEN RIGIDITY BALANCE SCORES AND REINTERPRETED RORSCHACH MALADJUSTMENT SCORES

Normal	Hysteric	Paranoid	Total Group
+.84 S.E. (.07)	+.46 S.E. (.21)	+.72 S.E. (.13)	+.60 S.E. (.09)

1. They imply that the concept of a general trait or general factor of personality rigidity is a fictional one. Although there appear to be certain *general* personality rigidity *trends,* to conceive of them in any simple unidimensional sense or even in a moderately complex *trait sense* is an oversimplification of the facts.

2. They point to the probability that personality rigidity varies at different levels of the personality structure.

3. They imply that "good" or healthy adjustment is usually marked by certain optimum relationships among the rigidity levels in the personality structure.

4. They imply that factors which lead to a disturbance in present adjustment, which require a difficult shift in rigidity defenses, result in a breakdown of optimum relations among the various rigidity levels.

5. Lastly, they indicate that in extreme cases where conflict and disturbance in defenses give way to a markedly regressive, settled neurotic adjustment or regressive-like psychotic adjustment, the various rigidity levels tend to resume the optimum relations characteristic of "non-disturbed" adjustment.

E. ANALYSIS OF INDIVIDUAL RIGIDITY PATTERNS

In this section an attempt will be made to analyse the various kinds of rigidity patterns that appear in each of the groups and to relate such patterns to degree of maladjustment, clinical symptomatology, and the kind of description of himself the subject gives on the STDCR.

One of the chief problems in analysing the rigidity patterns of a range of subjects is that the large number of rigidity scores which must be considered simultaneously in each case are difficult to conceive as a unity. The underlying pattern is concealed by the fact of complexity. In order to deal with this confusing complexity a number profile system has been worked out which gives a simplified picture of rigidity patterns. This system will now be briefly described. Four variables are considered in the course of picturing a particular rigidity pattern:

1. The degree of present *disturbance* in the pattern—i.e., its deviation from the "ideal" profile previously described in terms of the rigidity balance score.
2. *Generalized trends* toward rigidity—as indicated by the total rigidity score.
3. Degree of rigidity at level I as indicated by a representative score from level I—viz., the independent TAT interpretation score.
4. Indications of unusually low rigidity when dealing with "easy" or "peripheral" rigidity tasks—as measured by the number of certain level II and III tasks for which unusually low rigidity scores are obtained (see section D above).

These rigidity variables may be converted into numbers (as will be described below) and the rigidity pattern represented by a numerical profile in which a particular position is assigned to each variable. Thus: (see top of next page). In this way, by acquainting oneself with the significance of each position in the profile, one may grasp the general picture of an individual's rigidity pattern.

The basis for assigning numerical values to the various positions in the numerical profile is indicated by the

Position One	*Position Two*	*Position Three*	*Position Four*
(degree of disturbance)	(general rigidity trends)	(level I rigidity)	(indicator of unusually low peripheral rigidity)

criteria set forth in Table 15A.

In general, then, if one reads from left to right in any rigidity numerical profile one may note degree of pattern disturbance, general rigidity trends, degree of rigidity under emotional stress, and indications of degree of unusual peripheral flexibility.

Table 16 shows the total range of rigidity patterns that one might theoretically obtain in terms of the four rigidity variables in the profile. It can be seen that the actually obtained patterns tend to scatter over the whole theoretical range. However, one finds instances (e.g., 1120 through 1131) where there are sudden gaps in the distribution. One can only speculate about the meaning of such gaps. It is possible that certain rigidity pattern characteristics are psychologically mutually exclusive.

The patterns for any particular diagnostic group cluster moderately, but there is much overlap among the groups. As will be demonstrated in the analysis to follow, a rigidity profile does not have an absolute significance. Rather, it must be interpreted *relative to other variables,* such as clinical status, symptomatology, and degree of maladjustment. When rigidity profiles are interpreted in a *relative* sense, their distributions *within* the diagnostic groups reveal themselves to be psychologically meaningful and separable into sub-groupings that overlap only very moderately. An analysis of the rigidity pattern distributions in the various diagnostic groups follows.

1. *Patterns in the Normal Group*

a. Rorschach test. If each subject's rigidity pattern is placed on a continuum relative to the *rank position* of the subject's Rorschach maladjustment score (as shown in Table 17), some very interesting cluster effects appear. That is, one perceives that the rigidity profiles within certain ranges on the Rorschach maladjustment continuum show similarities to each other and are different from the profiles found at other points on the continuum. By inspection one is able to isolate four rather distinct groupings or clusters of profiles. The first of these groups consists of the four subjects with among the

TABLE 15A

POSSIBLE VALUES FOR THE POSITIONS IN THE RIGIDITY PATTERN PROFILE

Position One (Pattern Disturbance)			Position Two (Generalized Rigidity Trends)			Position Three (Level I Rigidity)			Position Four (Unusual Peripheral Flexibility)		
Degree of Disturbance	Ranks	Value	Degree of Rigidity	Rank	Value	Degree of Rigidity	Ranks	Value	Low Rigidity Peripheral Tasks	Ranks	Value
High	1–22	3	Very high	1–10	4	TAT high	1–20	3	Color	1–10	N[a]
Moderate	23–40	2	High	11–20	3	Moderate	21–40	2	Annoyances (S)	1–15	
Low	41–60	1	Moderate	21–45	2	Low	41–60	1	Annoyances (O)	1–10	
			Low	46–60	1				Trait Judgment	1–15	
									Aspiration[b] (S)		
									Self[c] (Unusual Series)		
									Self[c] (Ordinary Series)		

[a] "N" indicates the number of instances in which the subject displayed unusually high peripheral flexibility.

[b] The scoring here is more complicated. If the aspiration F rank is below twenty and the aspiration S rank is nine or more ranks higher, this represents an instance of unusual peripheral flexibility.

[c] Here the scoring of easy peripheral flexibility is based on a comparison of the flexibility shown for a task easily controlled by the subject as contrasted with a task not easily controlled. If the self rank is more than five ranks higher than the independent TAT interpretation rank, it is considered that an unnecessary peripheral flexibility has been mobilized.

TABLE 16

Total Possible Range of Rigidity Profile Patterns, and Those Actually Appearing in the Three Experimental Groups

Profile	N[a]	H	P	Profile	N	H	P	Profile	N	H	P
1 1 1 0	3	1	2	2 1 1 0				3 1 1 0			
1 1 1 1[b]	3	1		2 1 1 1	2			3 1 1 1	2		
1 1 2 0				2 1 2 0				3 1 2 0			
1 1 2 1				2 1 2 1				3 1 2 1		1	
1 1 3 0				2 1 3 0				3 1 3 0			
1 1 3 1				2 1 3 1				3 1 3 1			
1 2 1 0	1			2 2 1 0				3 2 1 0			
1 2 1 1	1			2 2 1 1				3 2 1 1		2	
1 2 2 0	2			2 2 2 0		1	1	3 2 2 0			
1 2 2 1				2 2 2 1	2			3 2 2 1	2		3
1 2 3 0		1		2 2 3 0				3 2 3 0			
1 2 3 1			1	2 2 3 1				3 2 3 1			4
1 3 1 0				2 3 1 0				3 3 1 0		2	2
1 3 1 1				2 3 1 1	1			3 3 1 1			1
1 3 2 0				2 3 2 0				3 3 2 0			
1 3 2 1		1		2 3 2 1				3 3 2 1			1
1 3 3 0				2 3 3 0		1		3 3 3 0			
1 3 3 1				2 3 3 1		1	2	3 3 3 1		2	
1 4 1 0				2 4 1 0				3 4 1 0			
1 4 1 1				2 4 1 1				3 4 1 1			
1 4 2 0	1			2 4 2 0		1		3 4 2 0			
1 4 2 1				2 4 2 1		1	1	3 4 2 1			
1 4 3 0		1		2 4 2 0		1		3 4 3 0			
1 4 3 1				2 4 3 1		1	2	3 4 3 1		1	

[a] N = number of normal subjects manifesting the indicated pattern; H = number of hysterics; P = number of paranoids.

[b] The number "1" here indicates that there may be one or *more* scores that imply unusually low peripheral rigidity; that is to say, in this table distinction is made only between the *presence* (indicated by "1") or *absence* (indicated by "0") of unusual peripheral flexibility.

TABLE 17

Rigidity Patterns of Normal Subjects Ranked According to Rorschach Maladjustment Scores and Separated into "Cluster" Groups

Group I	Group II	Group III	Group IV
Subject 7 1 1 1 0	Subject 12 1 2 2 0	Subject 4 2 1 1 2	Subject 13 3 1 1 2
Subject 19 1 2 1 1	Subject 20 1 2 2 0	Subject 10 2 3 1 1	Subject 17 3 1 1 2
Subject 14 1 1 1 0	Subject 1 1 4 2 0	Subject 11 1 1 1 1	Subject 3 3 2 2 5
Subject 8 1 1 1 0	Subject 2 (not classified) 1 2 1 0	Subject 15 2 1 1 2	Subject 5 3 2 2 3
		Subject 16 1 1 1 3	
		Subject 6 2 2 2 3	
		Subject 18 2 2 2 1	
		Subject 9 1 1 1 2	

lowest Rorschach maladjustment scores. With but one minor variation their common rigidity pattern takes this form: 1 1 1 0. In terms of what has previously been said about the general significance of each of the profile positions, one may conclude that this 1 1 1 0 profile is marked by low *emotional* rigidity; an absence of any generalized rigidity trends; and freedom from disturbance in the interrelations of the rigidity levels. Furthermore, one is not able to detect any erratic manifestations of unusually low *peripheral rigidity*. The whole pattern conveys a picture of optimum distribution of rigidity defenses within the personality structure. Stated in more positive terms, the individuals in this group seem to show the following rigidity characteristics:

1. They apparently seem not to feel threatened by most situations—as is indicated by the fact that they have no need for maintaining any sort of persistent anticipatory guarding or rigidity set.

2. They seem to maintain their composure under emotional stress very well. In unusually demanding situations they mobilize quickly; for instead of retreating behind rigidity defenses, they mobilize increased flexibility for the emergency.

3. Although they are capable of freeing themselves from rigidity tendencies to a high degree when the occasion requires, they tend to maintain a certain basic minimum peripheral rigidity that is free of variations in the direction of extreme flexibility.

The second pattern group[2] consists of three individuals who are, in general, moderately higher in Rorschach maladjustment than the individuals of *group one*. Their basic pattern takes this general form: 1 2 2 0. It is obviously a more guarded and rigid group than *group one*. Level I rigidity is moderately high and *general* rigidity trends vary from moderate to extreme. However, there is a complete absence of disturbance in rigidity level relations and no signs of unusually low variations in peripheral rigidity. This group is especially interesting because it consists of individuals who are making a healthy adjustment, but in terms of much more rigid defenses than group one. The subjects in this group tend to approach situations with greater anticipatory rigidity and to react to stress demands in a more guarded fashion. However, they *consistently* maintain an optimum balance between peripheral and emotional level rigidity and are capable of increasing their customary level of flexibility when more important emotional situations are in question. This balance and ability to increase flexibility under stress are apparently the most important factors in the group's healthy adjustment; for as will be shown, many more maladjusted individuals display greater freedom from rigidity—but not the same degree of balance in the relations among the rigidity levels of the personality structure.

The eight subjects in group *three* show considerable variation in their pattern profiles. One finds a range from 1 1 1 1 to 2 3 1 1. However, they are in general distinguished from those in the groups preceding them by their consistent manifestation of unusually low peripheral rigidity in position IV and frequently also by signs of moderate rigidity balance disturbance in position I. Persons in this group seem to have some difficulty with their present defenses. Although they show little generalized anticipatory guarding and little need to retreat to a more rigid position when faced with stress demands, they do not manifest the same coordinated unity in rigidity organization that is characteristic of the healthier normals in the other groups. They assign a greater importance to peripheral demands and irregularly they meet such demands with inappropriate mobilizations of flexibility which in healthier subjects are usually reserved for facing more important (emotional) situations.

Since the phenomenon of unusually high flexibility mobilization in reaction to peripheral demands is rather prominent in this group it would be well to consider briefly its meaning at this point. The phenomenon refers specifically to such behavior as claiming an unusually high number of self interests; denying that more than a very few things are annoying; acceptance of an extreme number of pictured persons (especially in the unpleasant series) as being like the self; showing an *overwillingness* to raise aspirations during success; etc. These peripheral mobilizations may be thought of as serving two functions:

1. On the one hand, some of them (e.g., over-responsive raising of aspirations during success) seem to represent an attempt to secure easy satisfaction and success. They represent, as it were, an attempt to take a peripheral activity out of its peripheral context and its conversion into a central source of satisfaction.

2. On the other hand, certain of the peripheral mobilizations (e.g., acceptance of an unusual number or unpleasantly pictured persons as being like the self) seem to represent an over-responsiveness to tasks which are interpreted as being much more threatening than most other subjects see them. That is, what

---

[2] Subject 2 is omitted from any of the pattern classifications because her pattern shows few similarities to any of the patterns in its vicinity on the maladjustment continuum.

remains a fairly peripheral type of threat for most subjects become an important threat for others and these latter deny their vulnerability by showing an extreme overwillingness, for example, to admit the possibility that all sorts of unpleasant (even psychotic-looking) individuals are like themselves. In general, peripheral mobilizations of this sort may be thought of as representing an over-sensitivity to peripheral threat.

The peripheral mobilizations in *group three* included both types; but the use of peripheral activities for *easy satisfaction* predominates.

The *fourth group* is one that contains the four most maladjusted normal subjects, according to the Rooschach test. The profiles in the group all have the same value in position I, values above 0 in position IV, and approximately similar values in positions II and III. Most clearly, they have in common the fact of their extremely high disturbance in rigidity balance relations and their frequent unusual peripheral flexibility mobilizations. Actually, they are only more extreme examples of group three. They have even greater difficulty in coordinating their defenses and a more pronounced concern with peripheral demands. It is important to note that the two *most maladjusted* subjects in this group show, in addition, a tendency toward moderate *generalized rigidity trends.* These two individuals are guarded and unable to increase their flexibility to meet efficiently situations with important emotional significance.

b. STDCR. An attempt was made to relate STDCR scores to the above described groupings —with the following results:

Group one.—All of the individuals in this group give STDCR scores which attain or moderately exceed in "goodness" *general* population means. Three of the four subjects attain or exceed the middle rank position for the normal *experimental* group. Furthermore, the same three subjects give STDCR *variation scores* that are in the upper twenty per cent of the normal group. That is, the well integrated healthy individuals of *group one* describe themselves in favorable or "good" terms. However, it is not an extremely "good" self description that they give and it is not a description of extreme uniformity. Their evaluations of factors STDCR vary in terms of the particular factor concerned.

Group two.—All of the individuals in group two give STDCR scores that cluster either at the mean or slightly below the mean of general population norms. However, all of their STDCR *sums* are *below* the middle rank position of the normal experimental group; and with but one exception their *STDCR variation scores* are distinctly above the middle rank position of the normal experimental group.

Clearly, these subjects tend to give a relatively more unfavorable description of themselves—although they show the same ability as the subjects in group one to vary their evaluations from factor to factor. The fact that this group attempts to give only a relatively restrained picture of itself ties in well with the fact that its rigidity pattern is definitely a guarded restrained one. Apparently, the general style of defense reflected in the rigidity pattern reflects itself also in the kind of description of himself the subject presents.

Group Three.—The STDCR sums here scatter widely and follow no aparent uniformity. Five of the eight STDCR *variation scores* fall below the middle rank position and suggest that the individuals in this group tend to be less discriminating than those in groups one and two relative to the single STDCR factors. That is, they seem to describe themselves in more generalized "halo" terms.

Group Four.—The STDCR scores of this group are quite uniformly *above* the general population means. Three of the four subjects have *STDCR sums* which are distinctly above the middle rank position of the normal experimental group. Furthermore these same three subjects' STDCR *variation* scores are in the lowest twenty per cent of the normal group. Generally, it would apear that *group four* individuals give a very one-sided "good" descriptions of themselves. They do not discriminate well relative to the single STDCR factors. It may be postulated that such biased "favorable" self descriptions perhaps represent an attempt to conceal or deny the disturbance which is so apparent in both the rigidity pattern and in the Rorschach material.

Although one can speak only in terms of trends, there are indications from the above that when STDCR scores are interpreted relative to rigidity patterns they take on a functional meaning. That is, one can understand the kind of description of himself the subject gives as a reflection of specific relationships and signs of disturbance indicated in his rigidity pattern.

2. *Patterns in the Hysteric Group*

The hysteric group may also be analyzed into typical patterns. Although pattern clusters in the normal group were verified entirely by similarities in Rorschach maladjustment scores, an *additional* variable (viz., clinical symptoms) will be used to aid such verification in the hysteric group. Each identified pattern will be described in turn.[3]

---

[3] Subject 32 is not included in any of the groups to be described below because there are data which tend to contradict a diagnosis of

TABLE 18

INDIVIDUAL *STDCR* SCORES, RANKS OF *STDCR* TOTAL SCORES, AND RANKS OF *STDCR* VARIATION SCORES FOR THE NORMAL SUBJECTS IN EACH RIGIDITY GROUP

Subjects	S[a]	T	D	C	R	Rank of Sum of STDCR Scores[b]	Rank of Variation Scores[c]
Rigidity Pattern Group One							
Subject 7	6	10	9	5	8	5.0	1.5
S 19	7	7	6	4	8	10.5	5.5
S 14	5	7	7	6	5	14.0	15.5
S 8	5	10	6	5	7	8.0	5.5
Rigidity Group Two							
S 12	8	3	6	4	8	16.0	1.5
S 20	6	2	5	4	5	19.0	5.5
S 1	5	3	5	4	6	18.0	9.5
Rigidity Group Three							
S 4	6	9	7	5	5	10.5	8.0
S 10	7	5	8	6	6	10.5	11.5
S 11	6	2	3	2	7	20.0	5.5
S 15	6	6	7	5	6	14.0	15.5
S 16	7	9	9	7	8	2.5	15.5
S 6	8	7	9	7	6	6.0	11.5
S 18	6	5	7	6	6	14.0	15.5
S 9	8	7	9	6	9	4.0	9.5
Rigidity Group Four							
S 13	8	10	9	8	8	1.0	19.5
S 17	9	7	9	8	7	2.5	15.5
S 3	6	6	8	6	8	7.0	19.5
S 5	6	4	9	9	4	10.5	3.0

[a] Each factor is represented by a C-score based on adults.
[b] Highest rank for highest sum.
[c] Highest rank for greatest variation.

Group one.—This group consists of five subjects who are among those hysterics showing least maladjustment on the Rorschach test. Clinically they all have similar symptoms. Thus, the chief complaint of each subject revolves about *localized* pain or disturbance in some part of the body, (e.g., heart palpitation, pain in stomach, pain in back, pain in legs). In this group both positions II and III (with one exception) are 3 or higher. Aside from case 33, position IV is consistently O. Also the values in position I do not exceed 2. The members of the group tend to be rigid at all levels. Their personality structures are protected by an elaborate series of high rigidity barriers. They show little capacity for mobilizing flexibility to meet stress situations having emotional significance. In general, they are constricted persons who seem to depend upon guarded rigidity to evade dealing with problems. However, in none of these cases does the degree of disturbance in pattern consistency (as shown by the balance score) exceed a moderate level. This implies that although the pattern is a tight rigid one, it is still rather stable and does permit the subject to function moderately well. Furthermore, it may be noted that mobilizations of unusual peripheral flexibility which are associated with special disturbance in rigidity defenses are also absent in this group.

Group two.—Group two consists of seven sub-

---

conversion-hysteria. First of all, her clinical symptoms consist chiefly of rather vague complaints of insomnia and some slight difficulty in moving one hand. Secondly, her Rorschach maladjustment score is "better" than the mean normal score. Finally, her rigidity pattern is that of a group one normal. Subjects 23 and 25 are also not included in any of the groupings. They simply do not fall into any of the patterns that can be identified. This may possibly be attributed to the limited sampling of conversion hysterics represented in the group.

TABLE 19

RIGIDITY PATTERNS OF CONVERSION HYSTERICS RANKED ACCORDING TO RORSCHACH MALADJUSTMENT SCORES AND SEPARATED INTO "CLUSTER" GROUPS

Group I	Group II	Group III
Subject 32 (not classified) 1 1 1 0	Subject 40 3 2 3 1	Subject 31 3 2 1 2
Subject 26 2 3 3 0	Subject 22 3 4 3 3	Subject 37 1 1 1 2
Subject 30 1 4 3 0	Subject 38 2 4 3 1	Subject 23 (not classified) 2 4 2 0
Subject 33 2 3 3 1	Subject 28 2 4 2 1	Subject 21 3 1 2 2
Subject 29 2 4 3 0	Subject 35 3 2 3 2	Subject 34 2 2 2 0
Subject 39 1 2 3 0	Subject 27 3 3 3 2	Subject 24 1 3 2 2
Subject 25 (not classified) 3 2 1 2	Subject 36 3 3 3 2	

jects whose Rorschach maladjustment scores cluster from rank five to rank eleven (rank one indicating the most maladjusted). In contrast with group one, it embraces those who manifest more serious and generalized symptoms (e.g., generalized body pains, persisting vomiting, and dizzy spells). It includes what might be called the larger "middle group" of hysterical subjects. The patterns typical of the group are marked by values of 2 or higher in position I, a value of 3 (with one exception) in position III, and a value of 1 or higher in position IV. These patterns are all similar in that they picture individuals who manifest strong rigidity trends, but whose rigidity defenses are in a state of high disturbance. The group two hysterics show about as much and sometimes less rigidity than the healthier hysterics; but they lack the same coordinated unity. Their rigidity is less consistent (in terms of balance). On the one hand, they are even less capable of mobilizing increased flexibility to face *important* demands; and on the other hand they are extremely more prone to bursts of non-functional peripheral flexibility. This is another indication of how much more important the organization factor is than mere highness or lowness of general rigidity trends.

Group three.—Group three includes five subjects:

1. Four hysterics with extreme maladjustment scores who show very severe clinical symptoms. Three of these subjects' symptoms center about tremors or convulsions involving the entire body. One subject has had periods of mutism; feels extremely nervous; and manifests *tremor* and numbness in one arm.

2. Group three includes also one subject whose Rorschach score places her *below* the mean of the hysteric group in maladjustment. However, despite her moderately good showing on the Rorschach test, she has manifested such symptoms as mutism, paralysis of lower limbs, brief convulsive spells, and periods of amnesia. The value of the rigidity pattern is particularly emphasized in this instance; for it classifies this subject with the other hysterics showing most serious symptoms—even though the Rorschach fails to do so. The rigidity patterns typical of the group are most characterized by position III values which do not fall below 2 and by position II values (with two exceptions) which also do not fall below 2. Clearly, these hysterics are much less rigid than those that have already been described. They are less protectively guarded, less constricted by general rigidity barriers; and do not show the same tendency to retreat to rigidity when exposed to emotional stress demands. However, all of their rigidity patterns give evidence of considerable internal disturbance in terms of balance scores or inappropriate peripheral flexibility. The fact that this group manifests the poorest maladjustment scores and the most extreme symptoms, taken in conjunction with the fact that it is the least

TABLE 20

Individual *STDCR* Scores, Ranks of *STDCR* Total Scores, and Ranks of *STDCR* Variation Scores for the Conversion Hysteric Subjects in Each Rigidity Group

Subjects	S	T	D	C	R[a]	Rank of Sum of STDCR Scores[b]	Rank of Variation Scores[c]
Rigidity Pattern Group One							
Subject 26	1	2	0	0	4	20.0	10.0
S 30	4	8	6	5	7	3.5	8.0
S 33	3	5	5	5	4	8.5	14.5
S 29	5	5	4	3	5	8.5	14.5
S 39	3	7	2	3	1	15.5	2.5
Rigidity Group Two							
Subject 40	5	10	7	6	6	1.0	6.0
S 22	4	7	5	5	5	6.0	12.0
S 38	3	4	2	2	5	15.5	12.0
S 28	4	6	5	5	1	10.0	2.5
S 35	4	6	4	3	2	12.5	8.0
S 27	6	6	4	3	8	5.0	4.5
S 36	1	3	1	2	3	19.0	16.5
Rigidity Group Three							
Subject 31	5	5	3	3	4	11.0	16.5
S 37	5	3	3	3	5	12.5	19.0
S 21	3	2	2	2	4	17.0	19.0
S 34	5	6	4	4	4	7.0	19.0
S 24	1	5	3	3	0	18.0	4.5

[a] Each factor is represented by a C-score based on adults.

[b] Highest rank for highest sum.

[c] Highest rank for greatest variation.

rigid group of hysterics, leads one to postulate the following: As hysterics become sicker and more maladjusted, they are less able to meet their problems in terms of defenses based on evasive rigidity and these increasingly inadequate defenses begin to break down into a looser less organized form.

The relations of STDCR scores to the hysteric rigidity pattern groups are not so clear as they were relative to the normal subjects. Most of the trends that occur are most apparent when groups *one and two* are contrasted with group *three.* The hysterics of groups one and two show no uniformity in the kinds of descriptions they give of themselves; but the hysterics of group *three* give a very unfavorable description of themselves. Four out of five in this group obtain total STDCR scores below the middle rank position of the hysterics and they all tend to give individual factor scores that fall below the mean of the *general* population. Furthermore, while eleven of the twelve hysterics in groups one and two show at least moderate *variation* (ranks 1-14) in the range of their individual STDCR factor scores; four of the five *group three* hysterics show very low variation (ranks 16-19) in their STDCR scores. That is, not only do the *group three* hysterics give an unfavorable description of themselves, but also they give the description in a gross generalized fashion that does not show discrimination for the single STDCR factors. These STDCR differences are logical in terms of the rigidity differences of the contrasting groups. That is, groups *one and two* are marked by their guardedness, their retreat to evasive rigidity when faced with problems. Contrastingly, group *three* shows much less guarding and is characterized by a loosening of rigidity defenses. Thus, the less guarded group tends to describe emphatically and openly its subjective unhappiness; while the more guarded evasive group is less able to do this. Of course, another point to consider is that the group *one and two* hysterics are not as ill and maladjusted as the group *three* hysterics; and so perhaps do not feel as subjectively disturbed as the latter group. However, an explanation based purely on degree of maladjustment would not provide an understanding of why certain specific *group one and two* hysterics with the most extremely rigid patterns (but not the very best maladjustment scores) give descriptions of themselves which exceed in "goodness" the general population means.

TABLE 21

RIGIDITY PATTERNS OF PARANOID SCHIZOPHRENICS RANKED ACCORDING TO RORSHACH MALADJUSTMENT SCORES AND SEPARATED INTO "CLUSTER" GROUPS

Group I	*Group II*
Subject 55 3 2 2 1	Subject 59 3 3 1 0
Subject 53 3 3 2 2	Subject 42 1 1 1 0
Subject 43 2 2 2 0	Subject 48 3 3 1 0
Subject 60 3 2 3 2	Subject 47 1 1 1 0
Subject 44 2 3 3 2	Subject 46 3 3 1 4
Subject 57 2 3 3 1	
Subject 50 3 2 3 3	
Subject 56 1 2 3 2	
Subject 41 3 2 3 2	
Subject 49 2 4 2 1	
Subject 54 2 4 3 2	
Subject 52 2 4 3 1	
Subject 45 3 2 2 3	
Subject 58 3 2 3 5	
Subject 51 3 2 2 3	

3. *Patterns in the Paranoid Group*

The paranoid rigidity patterns cluster along the Rorschach maladjustment continuum into two very distinct groups. Within these groups there are certain subsidiary patterns which will also be noted.

Group one.—The profiles for this group are characterized by values of 2 or more in positions II and III and values of 1 or more (with a single exception) in position IV. Group one includes the fifteen subjects who have the lowest Rorschach maladjustment scores among the paranoids. These subjects tend to manifest relatively moderate to extremely high rigidity trends. In almost all instances they show themselves to be guarded and unable to mobilize increased flexibility when faced with stress or emotional demands. Their rigidity patterns correspond to the concept of the "rigid paranoid." Marked disturbance in the organization and balance of the pattern is clear in almost every instance. This disturbance shows itself either in terms of poor rigidity pattern balance scores or in the form of inappropriate peripheral flexibility.

Certain interesting pattern variations appear at the two *maladjustment extremes* of the cases in group one. Thus, the three *least sick* individuals in the group are associated with these patterns: 3 2 2 1, 3 3 2 2, and 2 2 2 0. They are less completely rigid and less subject to peripheral flexibility mobilizations. They do not retreat so strongly behind rigidity defenses when it is necessary to deal with a stress situation. At the other more maladjusted extreme of *group one* are also found individuals whose rigidity patterns tend to be less guarded and constricted then the middle segment of the cluster; but they, in contrast, show significantly more disturbance in balance scores and more peripheral irregularity than the least sick paranoids. In the one instance decreased rigidity implies less pathology and in the other it seems to imply more pathology.

Group two.—This group consists of the five paranoids who show the highest degree of maladjustment and looseness on the Rorschach test. The characteristic patterns have in common values of 1 in position III and (with one exception) values of 0 in position IV. These paranoids display much less rigidity than one finds associated with the other paranoids. Several do manifest rather high rigidity trends when dealing with situations that have only *peripheral* importance to them, but they make little attempt to guard themselves in *stress* situations by using rigidity barriers. Indeed, relative to materials with emotional significance their flexibility is so extreme as to warrant the use of the term "loose." In contrast to the *group one* paranoids who, it may be postulated, are still seeking to defend themselves against their chief problems by a constriction process, these *group two* paranoids have tended to drop their defenses and seem to have given up trying to handle their problems actively. Of course, this last supposition applies in different degrees. Thus, two of the cases (subjects 42 and 47), do show an almost complete absence of rigidity at any level and furthermore manifest few signs of disturbance about the loose unrealistic level of adjustment to which they have regressed. But three subjects (59, 48, and 46) are able to maintain considerable *peripheral* rigidity when not subjected to stress demands; and furthermore their balance scores imply considerable disturbance.

Analysis of the STDCR scores in the two paranoid rigidity pattern groups indicates that with but a few exceptions the individuals in both of these groups tend to give scores higher ("better") than the general population means.

TABLE 22

Individual *STDCR* Scores, Ranks of *STDCR* Total Scores, and Ranks of *STDCR* Variation Scores for the Paranoid Schizophrenic Subjects in Each Rigidity Group

Subjects	S	T	D	C	R[a]	Rank of Sum of STDCR Scores[b]	Rank of Variation Scores[c]
Rigidity Pattern Group One							
Subject 55	3	9	8	10	2	7.5	2.0
S 53	4	3	3	3	4	18.0	19.5
S 43	6	4	6	6	4	13.5	14.0
S 60	6	7	8	7	5	6.0	12.0
S 44	5	6	4	3	8	13.5	4.0
S 57	8	9	8	8	5	2.0	6.0
S 50	5	8	9	9	6	3.0	6.0
S 56	5	10	8	8	0	9.0	1.0
S 41	2	5	3	3	1	19.5	9.0
S 49	9	10	10	9	6	1.0	6.0
S 54	4	5	5	5	4	15.0	18.0
S 52	5	9	7	6	3	10.0	3.0
S 45	4	2	3	3	2	19.5	15.5
S 58	7	7	9	7	5	5.0	9.0
S 51	5	8	7	6	6	7.5	13.0
Rigidity Pattern Group Two							
S 59	5	6	6	5	5	12.0	19.5
S 42	7	8	8	6	7	4.0	15.5
S 48	4	7	3	4	4	16.0	11.0
S 47	5	3	3	3	4	17.0	17.0
S 46	6	4	8	7	4	11.0	9.0

[a] Each factor is represented by a C-score based on adults.
[b] Highest rank for highest sum.
[c] Highest rank for greatest variation.

That is, most of the paranoids give an at least moderately "good" descriptions of themselves. When applied to *group two* this fact might superficially seem to contradict what has been indicated about the looseness of defenses in that group; but if one assumes that a "good" self description represents only a peripheral kind of defensive effort, the apparent contradiction disappears. A comparison of groups *one and two* entirely in terms of the *paranoid* STDCR scores alone reveals that:

a. Nine of the fifteen subjects in *group one* have total STDCR scores which *equal* or *exceed* the middle rank position of the paranoid group; whereas

b. Four of the five subjects in *group two* have total STDCR scores *below* the middle rank position of the paranoid group.

c. Ten of the fifteen subjects in *group one exceed* the middle rank STDCR *variation* score of the paranoid group; whereas

d. Four of the five subjects in *group two* fall *below* the middle rank STDCR variation score of the paranoid group.

Apparently, the more rigid guarded paranoids not only *tend* to exceed the *general* population in the favorableness of their self descriptions but also to exceed the less rigid paranoids in this respect. However, despite their greater guardedness the *group one* paranoids show more variability (actually more selective discrimination) in their responses to the range of STDCR factors.

It may be pointed out that this difference in *variation* between the more and less maladjusted pattern groups holds true also for the normals and hysterics—perhaps implying some kind of rough relationship between the amount of discriminative variation shown in answering STDCR questions and degree of maladjustment.

4. *Summary of Section*

In general terms, the following has been shown in this section:

a. Each subject's rigidity scores may be expressed as a pattern which has a fairly precise meaning relative to such criteria as Rorschach maladjustment

TABLE 23

SUMMARY TABLE SHOWING FOR THE THREE GROUPS THE DISTRIBUTIONS OF RIGIDITY PATTERN PROFILES, RANKED ACCORDING TO CORRESPONDING RORSCHACH MALADJUSTMENT SCORES

Rank	Normal	Hysteric	Paranoid
1			3314
2			1110
3			3310
4			1110
5			3310
6			3223
7			3235
8		1322	
9		2220	
10		3122	
11			2431
12			3223
13			2432
14			2421
15			3232
16		3332	1232
17			
18		1112	
19	3223		
20		3332	
21			3233; 2331
22			2332
23			
24		3232; 2421	
25			
26		2431	
27			2220
28		3433	3232
29			
30		1230; 3231	
31			
32			
33		3212	
34			3322
35		2430	
36		2331	
37			3221
38	3225		
39	3112		
40			
41	1112; 3112		
42			
43			
44	2221; 2223		
45			
46	1113; 2112		
47			
48	1111; 2311		
49			
50	2112	1430	
51	1110		
52			
53	1420		
54	1220		
55		2330	
56	1220		
57	1110		
58			
59	1211		
60	1110		

scores, clinical symptoms, and STDCR scores.

b. Four rigidity patterns may be discerned in the normal group. These are listed in the order of their degree of maladjustment (from low to high): (1) A low rigidity group marked by a high degree of balance in level relations. (2) A more constricted group whose rigidity pattern is however characterized by balance and unified coordination. (3) A group showing few generalized rigidity trends, but characterized by moderate disturbance in rigidity *balance,* and tendencies toward inappropriate mobilizations of peripheral flexibility. (4) A group marked by extremely high disturbance in rigidity balance.

c. Three rigidity patterns may be detected in the hysteric group. These patterns are meaningful in terms of Rorschach maladjustment scores and also clinical symptoms. (1) A very constricted pattern group with only moderate rigidity balance disturbance. Clinical symptoms in this group are limited to localized pain. (2) A constricted group with an extremely high degree of rigidity balance disturbance and much unusual peripheral flexibility. Clinical symptoms are of a more general character. (3) A relatively less rigid, looser group which also manifests much disturbance in rigidity balance. Clinical symptoms are very extensive and serious (e.g., prolonged convulsive seizures).

d. Two rigidity patterns appear in the paranoid group. These are listed in their order of Rorschach maladjustment (low to high): (1) A highly constricted rigid group manifesting much disturbance in rigidity balance and a tendency toward inappropriate peripheral flexibility. (2) A loose group chiefly characterized by its low degree of defensive guarding.

e. The self description given by a subject on the Guilford STDCR test has a functional relationship to the pattern of his rigidity defenses. However, it is a relationship that varies with the clinical group in which it occurs. Thus, normal subjects whose rigidity defenses are in a state of disturbance and who manifest unusual peripheral flexibilities give a very favorable self description. In both the hysteric and paranoid groups the most favorable self description is given by those inidivaduals who are the best organized and whose rigidity defenses are in a state of tight constrictive guarding. Least favorable self descriptions are given by those whose rigidity is diminished and whose rigidity patterns tend toward looseness and "breakdown."

F. DIAGNOSTIC DIFFERENTIATION

The previous analysis has indicated that rather distinct sub-groups can be identified *within* each of the principal experimental groups. The question arises as to the possibility of using rigidity scores as an independent means of diagnosis—i.e., as a means of distinguishing normals from persons identified psychiatrically as sick and also of distinguishing persons in different psychiatric categories. Specifically, can rigidity scores be used to differentiate the normals from the hysterics and paranoids? Furthermore, can the hysterics be differentiated from the paranoids?

The rigidity balance score does not successfully give such differentiation. Even in terms of mean differences the normals are only moderately well separated from the other groups (although these differences are statistically significant) and the hysterics are not significantly differentiated from the paranoids.

In an attempt to work out a technique that would be more successful for diagnostic differentiation, several scoring systems were attempted. One such scoring technique based upon a series of penalty weights for all instances of unusually high or low peripheral rigidity, all instances of unusually low ability to mobilize flexibility for meeting stress situations, and all instances of unusual differences in rigidity for tasks at the same rigidity level proved to be moderately differentiating. It differentiated the normal subjects from the hysterics and the paranoids excellently. However, it failed to distinguish the hysterics from the paranoids successfully. That is, it was possible to separate those who seem to be making a normal adjustment from subjects who are clinically diagnosed as abnormal. However, since the primary purpose of this study is an analysis of rigidity phenomena and only very secondarily a consideration of diagnostic differentiation, the rather complicated weighting system evolved for the above is not described.

In general, the rigidity battery seems to have greatest potential usefulness in

TABLE 24

CONSTANTS OF THE DISTRIBUTIONS OF RIGIDITY BALANCE SCORES IN THE THREE EXPERIMENTAL GROUPS, AND DIFFERENCES IN BALANCE SCORES AMONG THESE GROUPS

Constants	N	H	P	N+H	N+P	H+P
M	+3.7	−8.4	−14.0			
σ	5.3	16.5	17.5			
σ_m	1.2	3.8	4.0			
D				12.1	17.7	5.6
σ_{diff}				3.9	4.1	5.5
C.R.				3.1	4.3	1.0

cases where other material (e.g., clinical behavior or psychological test results) identify an individual's illness rather definitely as being neurotic in contrast to psychotic—or vise versa. In such cases (as shown in the section of analysis of individual patterns) the rigidity scores are valuable for perceiving the kind of defenses an individual is relying on; how disturbed or unstable these defenses are; and how near they lie to the point of extremely serious disorganization. Thus, for example, one would have a better understanding of the seriousness and prognosis of a case of conversion hysteria if it were known whether the rigidity pattern was dominated by tight constriction, or unusual degree of peripheral flexibility, or perhaps tendencies toward looseness.

CHAPTER V

GENERAL DISCUSSION OF DATA

AS IN many other studies which attempt to get at some aspect of personality, the problem of specificity-generality has arisen prominently. That is, the question has been raised whether the rigidity phenomena here observed can possibly be explained either in terms of a "general rigidity" concept or in terms of specific, rather unrelated reactions to particular stimuli. The character of the results obtained imply that personality rigidity manifestations cannot accurately be described either in very specific terms or in very general terms. Neither single rigidity task scores nor gross summations of many scores have proved to be meaningful explanatory. In contradiction to such simple explanatory approaches, one is struck by the degree of complexity in the interrelations of the various rigidity measures. It is possible to note simultaneous correlated, contradictory, and complementary rigidity trends in the same subject. Indeed, aspects of the overt rigidity behavior of an individual may imply just the opposite of what is true with regard to that individual's basic rigidity pattern. For example, essentially rigid conversion hysterics or disturbed normals who are becoming increasingly rigid often display flashes of unusual flexibility when dealing with certain kinds of situations.

The data here obtained suggest that there is a reasonable basis for assuming that there are at least two roughly defined rigidity levels in any given personality structure—an *ego* level and a *peripheral* level. That is, if one observes the range of rigidity shown by individuals in meeting different kinds of situations, it becomes rather apparent that one can at least differentiate the degree of rigidity shown to situations that are not emotionally threatening from the degree of rigidity shown to those situations that do threaten the individual or somehow do call into question his self-esteem. The type of rigidity shown in reacting to the first kind of situation may be designated as *peripheral* or outer rigidity; whereas the second type of reaction may be called *ego* or "inner" rigidity.[1] If one postulates that the character of peripheral (outer) rigidity represents how much need the individual has to guard himself even in non-threatening situations and further assumes that ego (inner) rigidity represents the degree of defensive guarding an individual mobilizes when faced with emotional demands or threats, it may then be hypothesized that the pattern of relationships between these two levels gives us at least a vague picture of the manner in which a given individual's personality defenses are organized.

In these terms the defenses of a normal undisturbed person are marked by a series of balanced relationships between the inner level and the outer level. Their peripheral defenses usually manifest a certain basic minimum of rigidity; and the ego defenses show a consistent ability to mobilize a greater degree of flexibility than is habitually shown by the peripheral defenses. This relative relation is very important. There are several instances in which undisturbed normals demonstrate considerable rigidity at both levels; but these individuals are consistently able to call up a degree of flexibility for meeting stress situations that exceeds their customary outer rigidity. If such a normal person encounters problems or stresses that he cannot handle adequately with his current rigidity defenses, interesting adjustive changes begin to show up in his rigidity behavior. Disturbances in the peripheral defenses begin to appear. The balance of moderate to moderately high rigidity that previously characterized this level gives way to erratic fluctuations that vary from high *rigidity* to extremely high *flexibility.* Many situations that were previously adjusted to in terms of customary outer defenses can no longer be handled that way. The individual seems to be uncertain as to just how much he can count on his outer defenses and consequently many formerly "peripherally handled" situations become involved with ego defenses—with resultant erratic variations in the kind of rigidity shown at the peripheral level. As the individual's disturbance becomes even greater, his ego defenses become more seriously affected. They shift sharply in the direction of greater rigidity. They tighten and the ability to mobilize increased flexibility for meeting emergencies diminishes significantly.

---

[1] Although empirically the rigidity tasks were assumed to fall at three levels, it is not felt that the differences between levels II and III were marked enough to warrant a reliable conclusion as to the existence of defense levels intermediate between peripheral and inner defenses.

Passing from the normal to the conversion hysteric, one finds a rigidity defense pattern which is only a more extreme form of that seen in the very disturbed normal. However, in these hysteric cases rigidity at both defense levels is usually much greater, and the outer defenses are even more disturbed. But most outstanding of all is the fact of extremely increased ego-level constriction; the fact that stress demands seem to call forth only evasive retreat behind rigidity barriers. It is interesting to note that some few of the hysterics are characterized by a very rigid defense pattern and yet show only moderate peripheral disturbance and few signs of pattern instability. These are perhaps individuals who have worked out a neurotically constricted defense system capable of bearing with some efficiency the pressure of their current difficulties. In the most radically disturbed conversion hysterics one perceives a sharp shift in the pattern of rigidity defenses. An uncoordinated loosening of both the outer and inner defense levels appears. Disturbance persists in the form of erratic peripheral fluctuations and lack of balanced relationship between ego and peripheral barriers; and the individual no longer makes much attempt to deal with situations by retreating behind his rigidity defenses. At this stage of disturbance defenses dominantly based on *retreat to rigidity* are apparently no longer sufficient. Transition to a looser more regressive type of defense system seems to be in progress. This latter assumption is confirmed by the fact that regressive symptoms (e.g., generalized convulsive seizures) begin to appear in such cases.

Those paranoid schizophrenics who retain considerable control in their behavior and who are still strongly motivated to conceal and rationalize their psychotic thinking seem to make much use of rigidity in both their outer and inner defenses. They have a set to deal with situations in terms of minimum involvement. In general, however, their defenses tend not to be so completely closed and rigid as is the case with the middle group of *conversion hysterics.* There are more instances in which the ego defenses of these paranoids are not completely rigid barriers. But since their defenses are almost always in a state of high disturbance one can only interpret the lesser rigidity as a sign of greater vulnerability to disturbing stimuli. Paranoids who clinically (and in terms of psychological measures) seem to be sickest and most disorganized no longer make much genuine use of rigidity barriers. Although many of them have outer defenses which are very rigid, they do not seem to have any real inner defenses. This is demonstrated by the fact that demands which manage to evade their outer defenses or which are too stressful to be blocked peripherally, call forth very *loose,* obviously psychotic, reactions. Such reactions can rarely be called forth in the paranoid whose rigidity pattern indicates much rigidity at the ego level.

Altogether, this preliminary attempt to sketch in semi-precise terms the kinds of personality defense changes that accompany different degrees of personality maladjustment seems to give a meaningful consistent picture.

One very interesting problem that the above analysis of diagnostic groups brings to the fore has to do with the meaningfulness of our present diagnostic categories. Even if one grants that certain of the experimental subjects were incorrectly diagnosed, the fact remains that each of the experimental diagnostic groups includes *many* persons whose personality defenses seem to be organized in significantly different patterns. For example, among the conversion hysterics there are defense patterns marked by extreme rigidity at all levels and contrastingly there are patterns which are *loose* at all levels. The fact that these rigidity pattern differences are also frequently accompanied by radical differences in symptomatology leads one to think critically about the rationale of present classification schemes. At present, of course, psychiatric diagnosis is primarily a matter of evaluating symptomatology; but one may ask how valid such a procedure is if it leads to the grouping together of individuals whose personality defenses are apparently quite different. If future work with other fairly precise personality measures should confirm the kind of data here obtained about diagnostic groupings, serious criticisms might be raised against these groupings. It is obvious that important errors in treatment and understanding may result from assuming that a conversion hysteric "type A" is equivalent to a conversion hysteric "type B," or that a paranoid schizophrenic "type A" is equivalent to a paranoid schizophrenic "type B." The actual isolation of such types in a very exact sense could probably be achieved if a body of data were systematically accumulated about a range of personality defense phenomena, as has been done here relative to rigidity phenomena. The specific diagnostic value of the present approach to rigidity seems to lie in the fact that it can give valuable information relative to variations *within* diagnostic categories. Although it is true that when used alone the rigidity battery is at present able to distinguish only normals from abnormals, it is capable of giving very interesting insights into individual cases when used in conjunction with other materials (e.g., the Rorschach test). It may be added that the rigidity battery seems to have a high degree of sensitivity for detecting disturbance in normals. If further work with normals confirms this point, it is possible that an abbreviated form of the battery would prove of value in certain types of selection work.

Many of the points that have emerged

from the data have already been implicit or explicit in previous work. Thus, the whole viewpoint that has been expressed here with regard to *levels of rigidity* and the distinction between *outer* and *inner* rigidity has run through the thinking of psychologists and psychiatrists for several decades. It has been particularly marked in the thinking of the Lewinians, whose elaborate personality-structure scheme is prevaded with concepts like "sub-wholes of the personality structure," "ego barriers," and "peripheral barriers." Certain of the very specific findings of the present study (e.g., with regard to the rigidity of the conversion hysteric and the paranoid schizophrenic) have also been anticipated in previous work (Angyal, Shakow, and Rosenzweig). Actually, the value of the material presented in this paper lies not so much in its *specific* novelty, but rather in the fact that it represents at least an *initial* integration and confirmation of past isolated guesses and findings about rigidity behavior. Furthermore, it has value insofar as it has rationalized a particular set of measuring procedures that are based on a systematic point of view. Vague speculations about ego and peripheral defenses or changes that occur in these defenses under differing conditions are now subject to some kind of quantitative checks, even though they be only crude approximate checks.

In view of the importance that has been placed on the concepts of "ego" and "peripheral" rigidity in this study, it is pertinent that the conclusions reached here concerning them be related to previous work in the same area. The most extensive past experimental consideration of these concepts occurs in the series of "task-interruption" studies done by Rickers-Ovsiankina (17) and Bennett (5). Rickers-Ovsiankina found that schizophrenic subjects tended *not* to build up "completion tensions" relative to neutral tasks—as exemplified by the fact that they generally failed to resume such tasks following the removal of interposed interruptions. From such data she concluded that the outer peripheral boundaries of the schizophrenic are so fluid that tensions raised by neutral tasks are dissipated before they can form firmly segregated systems. Bennett then demonstrated that if schizophrenics are interrupted while performing tasks with real *emotional* significance to them, they manifest strong "completion tensions." She reasoned that this must imply that emotional materials have an opportunity to excite segregated tension systems because there are rigid barriers around the schizophrenic's ego area that prevent dissipation of tensions. That is, she demonstrated a real difference in the rigidity of ego boundaries as contrasted with peripheral boundaries. Furthermore, she (5, p. 241) noted that "in the central layers of the [*normal*] person, . . . communication [between systems of the *central* layers] seems to be even greater than that for peripheral systems." In a broad sense the results of the present study have been congruent with these interpretations based on task-interruption data. Thus, the ego-peripheral rigidity distinction has demonstrated itself clearly. Furthermore, it is particularly interesting that Bennett's conclusions regarding the lesser rigidity in *normal subjects* of ego areas as *contrasted with* peripheral areas has also been confirmed here.

However, there is one important point on which the present study stands in disagreement with the conclusions of Rickers-Ovsiankina and Bennett. This is relative to the first investigator's acceptance of the idea that the peripheral boundaries of schizophrenics are *in general* fluid and the second investigator's statement that the ego boundaries of schizophrenics are *in general* rigid. In terms of the complex variations in ego and peripheral rigidity found here in paranoid schizophrenics the above conclusions are misleading simplifications. Even within the limited range of a paranoid schizophrenic group one finds many cases of high peripheral rigidity, low peripheral rigidity, and also cases of marked ego fluidity and ego rigidity. As one reviews the data of Rickers-Ovsiankina and Bennett, one feels that a similar complexity actually characterized their results. However, it would appear that in the process of emphasizing and sharpening group comparisons they leaned toward excessive simplification. Their failure to underscore the complexity of rigidity phenomena may possibly also be attributed to the following: (1) The fact that their methods of measurement dealt with such a limited *sampling of behavior.* (2) The fact that they both assume that failure to develop completion tensions is due *only* to the absence of barriers necessary to prevent the dissipation of tensions. If one assumes that failure to develop completion tensions may be due also to failure of tasks to *excite motivation* by *penetrating* rigidity barriers, the possibility of a more complex interpretation of their data arises. That is, one should at least accept the *possibility* that failures to develop completion tensions may have alternative causes. On the one hand it may in some cases be due to high fluidity and in other cases it may be due to high rigidity.

Altogether, there seems to be justification for questioning the generalizations about peripheral and ego rigidity in schizophrenics that have been offered by Rickers-Ovsiankina and Bennett.

While working with the present rigidity battery certain problems and difficulties have become apparent which suggest several future research possibilities on the rigidity question.

1. First of all, it is clear that it would be desirable to work out a more balanced series of rigidity tasks—that is, a series in which ego-level tasks would not be so outnumbered by peripheral tasks, and in which at least one other defense level is identified between the peripheral and ego levels. Such refining of the rigidity battery would perhaps lead to more precise generalizations about rigidity relations, and possibly permit better discrimination of clinical groups.

2. Since this study was limited to women of three specified kinds of groups and of a particular range of intelligence, it is of course necessary to determine whether the results obtained apply to a wider range of subjects and how they are affected by extremes of intelligence.

3. Another interesting possibility would be a study of changes that occur in rigidity patterns when attempts are made to produce shifts in adjustment during different kinds of therapy. Thus, one might measure changes following electric shock. This approach might better our understanding of what kinds of defense reorganizations occur during such procedures. It would also be of special interest to work out the possible *prognostic* value of the rigidity measures.

CHAPTER VI

GENERAL SUMMARY OF FINDINGS

THE findings described below are based on a study of sixty women of average intelligence—twenty normal subjects, twenty conversion hysterics, and twenty paranoid schizophrenics. The chief points that emerge most clearly from the study may be summarized as follows:

1. No single rigidity measure based on any one task is successful either as a measure of "general" personality rigidity trends or of personality characteristics that might be thought of as related to rigidity.

2. Analysis of the relationships obtaining between total rigidity scores (a summation of many individual rigidity scores) and various other indices (e.g., Rorschach rigidity scores and clinical classification) of personality rigidity indicate that the total rigidity score is a moderately good measure of what might be called "general" rigidity trends. However, it has also been shown that the concept of general rigidity as a real entity is to a considerable degree fictitious and that one should probably not anticipate more than very moderate success in attempting to measure such a variable.

3. It has been shown that *in general* conversion hysterics and paranoid schizophrenics are more rigid than normal subjects. However, there are no clear differences between the conversion hysterics and paranoids.

4. It appears to be fruitful to conceive of rigidity behavior as correlated with a series of "defense levels" in the personality structure. The degree and character of the rigidity shown by an individual in dealing with a given situation seems to vary in proportion as that situation poses a threat or raises serious adjustment demands. In the normal personality certain optimum "ratios" are found to exist if one compares the amount of rigidity shown when dealing with non-threatening (peripheral) tasks as opposed to the amount of rigidity shown for threatening (ego involving) tasks. If the ratio or relation of ego level rigidity to peripheral level rigidity is for each individual expressed as a quantitative score indicating degree of departure from an ideal optimum ratio, correlations of such scores with certain measures of personality defense disturbance (e.g., the Rorschach and clinical symptoms) are significantly high.

5. Intelligence, in the formal psychometric sense, seems to have no clear relationship to the character of an individual's rigidity pattern—i.e., the relations obtaining between the rigidity of the peripheral defenses and ego defenses.

6. It is clear that personality rigidity may manifest itself in diverse, often unrecognized ways. Indeed, internal personality rigidity may show itself externally in a form that appears to be just the opposite of rigidity and even seems to imply extreme flexibility. Thus, numerous examples have been found in the data of subjects who are fearful and highly rigid in adjusting to ego involved situations; but who mobilize an unusual amount of unnecessary flexibility in adjusting to peripheral unimportant tasks. This non-functional and superfluous peripheral flexibility seems to represent a false "emergency reaction," a compensatory effort stimulated by the insecurity associated with an overly rigid ego defense system.

7. A "rigidity profile" may be con-

structed for each subject. This profile defines the nature of an individual's rigidity behavior in terms of four variables:

(*a*) The degree to which the individual's ratio of peripheral to ego rigidity departs from an ideal "normal ratio."

(*b*) The degree to which the individual's reactions seem to reflect a generalized rigidity that cuts across a wide range of response.

(*c*) The degree of rigidity shown when dealing with situations of a threatening or ego significance.

(*d*) The degree to which the individual seems to mobilize unusually great or unnecessary flexibility in meeting situations not usually conceived as ego involving.

It has been shown that within each of the experimental groups the "rigidity profiles" of subjects fall very logically into clusters. Those clustering together have approximately the same Rorschach maladjustment scores and display very marked similarities in clinical symptomatology.

8. The variety of "rigidity profiles" falling within the same diagnostic groups raises some doubt as to the rationale of present diagnostic classification schemes. Such questioning would have a more valid basis if confirmed by studies of other personality characteristics making use of at least moderately precise measures.

9. There are indications that the description which a subject gives of himself on the Guilford STDCR is functionally related to the kind of "rigidity profile" manifested by him. That is, those subjects whose "rigidity profiles" fall into the same clusters because of similarities in pattern tend also to give similar responses to questions about themselves on the STDCR. For example, all those normal subjects whose "rigidity profiles" are marked by unusual deviations from the ideal "normal ratio" of ego to peripheral rigidity and also by frequent unnecessary mobilizations of flexibility in the face of non-threatening tasks give very favorable descriptions of themselves. They may be contrasted with those normal subjects whose "rigidity profiles" show no such disturbance and who correlatively do not give extremely favorable or "good" STDCR descriptions of themselves. In general, the indications are that it is not possible to detect the clusters of subjects with similar STDCR scores or to explain the logic of such clusters without using the "rigidity profile" as a frame of reference.

10. It is possible in a statistically significant manner to differentiate the normal subjects from the abnormal subjects by use of various single rigidity indices (e.g., the total rigidity score or the ratio of peripheral to ego rigidity). However, this kind of differentiation is only on a rough group basis and not of sufficient clarity to permit the use of such measures as individual diagnostic tools in the conventional sense of the present diagnostic system of classification. The greatest "diagnostic" contribution of the rigidity battery seems to be toward detailed differentiation *within* clinical groups. Thus, by use of the "rigidity profile" it has been shown that one can with considerable accuracy discriminate disturbed normal subjects from well adjusted normal subjects (as defined by the Rorschach), conversion hysterics with limited localized symptoms from hysterics with more serious generalized symptoms, and well protected paranoid schizophrenics from more seriously disorganized paranoid schizophrenics.

APPENDIX I

VIGOTSKY SCORING

To score the Vigotsky sortings the following weighting system was used.

5 = A *basic* variation. This included all sortings in which the *total range* of blocks was divided into groups on the basis of *size* or *shape* or *color* or *number of sides*.

3 = For any variation in which the subject attempts to put one of each "kind" in each of the sorted groups. Thus, relative to form, this would be illustrated by a series of groups each containing a square, circular, and hexagonal block. Relative to color, it would be illustrated by a series of groups each containing five representative colors.

2 = A minor variation. Any repetition of a previous *basic* sorting—but with at least some minor change in the previous distribution of the blocks.

1 = Concrete variation. Any use of the blocks to make concrete objects or to build designs in jig saw fashion.

The above weighting system is based on clinical experience with the Vigotsky and is intended to assign each sorting to a level corresponding roughly to the complexity of analysis or abstraction involved.

APPENDIX II

TAT SCORING

The following weighting system was used to score TAT interpretations.

2 = A *basic*[1] interpretation which involves *all of the pictured* persons and revolves about a plot in which the *characters do something or are acted upon*.

2 = This weight also given for an interpretation which employs the same characters as a preceding *basic* interpretation—but which revolves about a plot in which the characters do something different or are acted upon differently.

1 = Given for an interpretation with different role takers than a preceding interpretation—but with the same plot or kind of action.

1 = Given for any basic interpretation which ignores any outstanding figure on the card.

0 = For mere card description or rejection of a card.

APPENDIX III

ASPIRATION SCORING

The following weighting system was used to evaluate aspiration changes that occurred following arbitrarily reported achievement scores. The table below is the basis for deriving three aspiration rigidity scores, as described in Chapter III, section B, 10.

Relation of Subject's Aspiration Shift to Shift in Arbitrarily Reported Performance Score	Weight
Aspiration shift is in desired direction and exceeds the arbitrary reported achievement shift by at least 2 points	+3
Shift[a] equals the arbitrary reported achievement shift	+2
Shift is less than the arbitrary reported achievement shift, but not less than one-half of the achievement shift	+1
Shift is less than one-half the arbitrary reported achievement shift	0
No shift in aspiration—when arbitrary reported achievement shift is only one point and subject's previous aspiration is more than two points distant from perfect achievement	—1
No shift in aspiration	—2
Shift in aspiration is *opposite* to arbitrary reported achievement shift[b]	—3

[a] If the arbitrary change reported to the subject is only one point greater than his previous aspiration and if the previous aspiration was more than two points distant from perfect achievement, a plus one is given instead of plus two.

[b] In evaluating the aspiration changes that occur with increasing failure only one-half of any positive credit (to nearest fraction) is given which might be earned on the *last* trial.

In a few cases where the character of the subject's expressed aspirations made it impossible to obtain four trials in the "success" series, the three obtained trials were equated to four by adding the average of the major trend (with fractions disregarded).

APPENDIX IV

SCORING FOR RORSCHACH RIGIDITY

Each of the weights below is a penalty for what is considered to be excess rigidity or restrictiveness. The larger the final summation of weights, the greater is the implied rigidity.

F per cent (F%)

(1) If the number of responses is 22 or fewer—

(*a*) An F% of 55-60 gives a score of 3.

(*b*) An F% of 61-70 gives a score of 4.

(*c*) An F% of 71-80 gives a score of 6.

(*d*) An F% of 81-90 gives a score of 8.

(*e*) An F% of 91-100 gives a score of 11.

(2) If the number of responses is over 22—

(*a*) An F% of 55-60 gives a score of 4.

(*b*) An F% of 61-70 gives a score of 5.

(*c*) An F% of 71-80 gives a score of 7.

(*d*) An F% of 81-90 gives a score of 9.

(*e*) An F% of 91-100 gives a score of 12.

Nunmber of Responses (R)

(1) If the number of responses lies between 0 and 15, inclusive, the score is 9.

(2) If the number of responses lies between 16 and 20, inclusive, the score is 6.

Percentage of Animal Responses (A)

(1) If the percentage of "animal" responses is 60-70, the score is 6.

(2) If the percentage of "animal" responses is 71-80, the score is 8.

(3) If the percentage of "animal" responses is 81-90, the score is 10.

Form Accuracy (F+)

(1) If the number of responses lies between 15 and 22, inclusive—

(*a*) And if the F% is at least 40 and not higher than 50—

An F+% of 85-90 gives a score of 5.
An F+% of 91-100 gives a score of 8.
(*b*) Or if the F% is 51 or higher—
An F+% of 85-90 gives a score of 6.
An F+% of 91-100 gives a score of 9.

If a record contains fewer than fifteen responses only one-half of any given weight applies.

(2) If the number of responses exceeds 22—

(*a*) And if the F% is at 35 and not higher than 50—

An F+% of 85-90 gives a score of 6.
An F+% of 91-100 gives a score of 10.
(*b*) Or if the F% is 51 or higher—
An F+% of 85-90 gives a score of 6.
An F+% of 91-100 gives a score of 12.

Card Turning

Less than two responses in which the card is viewed other than in the upright position gives a score of 4.

Color

Less than two FC responses is scored 6. (M.FC is not counted as an FC)

Whole Responses (W)

(1) If the percentage of W responses is 40-50—

(*a*) The score is 5 in records where the number of responses is 22 or under.

(*b*) The score is 6 in records where the number of responses exceed 22.

(2) If the W% is 51-60—

(*a*) The score is 7 where the number of responses is 22 or less.

(*b*) The score is 9 where the number of responses is over 22.

(3) If the W% is 61 or over—

(*a*) The score is 10 where the number of responses is 22 or less.

(*b*) The score is 12 where the number of responses is over 22.

Small Detail Responses (Dd)

(1) If the Dd% is 18-23—

(*a*) The score is 3 where the number of responses is under 22.

(*b*) The score is 4 where the number of responses is over 22.

(2) If the Dd% is 24-30—

(*a*) The score is 5 where the number of responses is 22 or under.

(*b*) The score is 6 where the number of responses is over 22.

(3) If the Dd% is 31-40—

(*a*) The score is 8 where the number of responses is 22 or under.

(*b*) The score is 9 where the number of responses is over 22.

(4) If the Dd% is 41 or over—

---

[1] For three of the cards (I, III, and VIII) a given basic variation or any other variation could earn only one credit. This was done because of the ease with which most subjects worked out interpretations for these cards that concealed rigidity tendencies on more "difficult" cards.

(*a*) The score is 10 where the number of responses is 22 or under.

(*b*) The score is 11 where the number of responses is over 22.

Movement Responses (M)

(1) If number of movement responses is less than two the score is 15.

(2) If the number of movement responses is less than one the score is 20.

Content

If four or more responses fall into the same content category (aside from "human" and "animal" responses)—

(*a*) And if the number of responses is 1 through 25 the score is 7.

(*b*) And if the number of responses is 26 through 40 the score is 6.

(*c*) And if the number of responses is 41 or over, the score is 5.

Average Time Of Initial Response (T/1R)

If average reaction time per initial response is

(*a*) 25 through 29 seconds, the score is 2.

(*b*) 30 through 40 seconds, the score is 5.

(*c*) Over 40 seconds, the score is 7.

Shading Responses (FY)

Less than two FY responses gives a score of 3.

APPENDIX V

SCORING FOR RORSCHACH MALADJUSTMENT

Each weight given below is a penalty for what is considered to be an indicator of personality maladjustment.

Small Detail Responses (Dd)

(1) If the Dd% is 18-25 inclusive, the score is 2.

(2) If the Dd% is over 25, the score is 8.

Whole Responses (W)

(1) If the W% is 30-35 inclusive—

(*a*) The score is 6 if the number of responses is 16-25.

(*b*) The score is 7 if the number of responses is 26 or higher.

(2) If the W% is 36-40 inclusive—

(*a*) The score is 8 if the number of responses is 16-25.

(*b*) The score is 10 if the number of responses is 26 or higher.

(3) If the W% is 41 or higher—

(*a*) The score is 12 if the number of responses is 17 or over.

(*b*) The score is 8 if the number of responses is 14-16.

(*c*) The score is 6 if the number of responses is 11-13.

(*d*) The score is 4 if the number of responses is 9-10.

Only one-half of any of the above penalty weights is given if more than one-half of the responses in the record are not minus or undifferentiated type of color responses (viz.,CF-and C).

Furthermore, except within the range 41% or higher, only one-half of a penalty weight is given if the number of responses in the record is less than 16.

(4) If the W% is less than 15, the score is 2.

(5) If the W% is less than 10, the score is 3.

(6) If one-fourth or more of the W's in a record lie beyond the middle of the response sequence of the particular cards on which they occur, the score is 3.

(7) If the sequence is confused, the score is 5.

Movement Responses (M)

(1) If the number of M's is less than 3—

(*a*) The score is 3 if the number of responses is 18 or less.

(*b*) The score is 5 if the number of responses is 19 or more.

(2) If the number of M's is less than 2—

(*a*) The score is 10 if the number of responses is 18 or less.

(*b*) The score is 12 if the number of responses is 19 or more.

(3) If the number of M's is less than 1—

(*a*) The score is 14 if the number of responses is 18 or less.

(*b*) The score is 16 if the number of responses is 19 or more.

(4) If an M does not occur on card III, the score is 6.

(5) One M of bad form quality (M minus) equals a score of 3.

(6) Two M minus responses equal a score of 10.

(7) Four M minus responses equal a score of 12.

(8) If 2-3 M's occur in small details, there is a score of 4.

(9) If 4 or more M's occur in small details, the score is 7.

(10) If there are over two M's in a record and if half or more are found beyond the middle of the cards on which they occured, the score is 5.

Loose Color or Color of Poor Quality (C, CF—, FC—)

(1) A pure color response (C) without any form is equal to a score of—

(*a*) 3 on card X.

(*b*) 2 on cards II or III.

(*c*) 2 on card VIII.

(*d*) 1 on card IX.

(2) A color response with some minimum form but of poor quality (CF—) and containing either blood content, fire content or anatomy content gives—

(*a*) A score of 3 on cards II, III, or IX.

(*b*) A score of 4 on cards VIII or X.

(3) Minus FC responses

(*a*) One FC minus response equals a score of 2.

(*b*) Two FC minus responses equal a score of 6.

(*c*) Three or more FC minus responses equal a score of 8.

Absence of Color

(1) If color response involving a good integration of form and color (FC+)

(*a*) Is absent on cards II and III, the score is 6.

(*b*) Is absent on cards VIII, IX, and X, the score is 9 (if FC given as *last* response on X, the score is 4).

(2) No color in the total record gives the score of 4.

(3) If there is only one FC in the total record and the rest of the color is C and CF minus, the score is 3.

Color Naming (CN)

(*a*) CN on one card equals a score of 12.

(*b*) CN on two or more cards equals a score of 16.

Form Quality (F+)

(1) If F+ is from 57 to 64 inclusive—

(*a*) An R of 15 through 20 (if less than 15 R, at least 9 of responses should be in the F category) will give a score of 5 for an F% 45-55; and a score of 7 for an F% of 56 and over.

(*b*) An R over 20 will give a score of
4 for an F% of 35-45.
6 for an F% of 46-55.
8 for an F% of 56 and over.

(2) If F+ from 50 through 56—

(*a*) An R of 13 through 20 will give a score of
7 for an F% of 35-40.
10 for an F% of 41-50.
12 for an F% of 56 and over.

(*b*) An R over 20 will give a score of
10 for an F% of 35-40.
15 for an F% of 41-50.
20 for an F% of 56 and over.

(3) If F+ is under 50—

(*a*) An R of 10 through 13 will give a score of
8 for an F% of 35-40.
10 for an F% of 4 -50.
15 for an F% of 51 and over.

(*b*) An R of 14 through 20 will give a score of
12 for an F% of 35-40.
18 for an F% of 41-50.
25 for an F% of 51 and over.

(*c*) An R over 20 will give a score of
18 for an F% of 35-40.
25 for an F% of 41-50.
30 for an F% of 51 and over.

Very Poor Form Quality

If a response is F minus, is a severe departure in form from the card area used, and contains very peculiar or fantastic content (e.g., "people catching blood in a cup," "dirty uterus," "inside of a womb"), that response is scored 5.

Number of Responses

(1) If there are less than 20 responses, the score is 2.

(2) If there are less than 15 responses, the score is 5.

(3) If there are less than 12 responses, the score is 9.

Average Time of Initial Responses (T/1R)

(1) If the T/1R is 25 through 35 seconds, the score is 2.

(2) If the T/1R is 36 seconds or greater, the score is 3.

Rejection of Cards

(1) Rejection of card I or IV or VI or IX equals a score of 7.

(2) Rejection of card II or III or V or VII or VIII or X equals a score of 9.

(3) Rejection of two cards in sequence results in an extra penalty of 5. (No more than two rejections, one for each of above categories, is counted).

Shading Responses

(1) 1 to 2 pure shading responses (Y) is scored 2.

(2) 3 to 4 pure Y responses is scored 4.

(3) 5 or more pure Y responses is scored 6.

(4) 2 to 3 shading responses with little form (YF) is scored 2.

(5) 4 or more YF responses is scored 5.

White Space Responses (S)

(1) 4 to 5 S responses is scored 2.

(2) 6 or more S responses is scored 4.

Popular Responses (P)

(If less than 12 R in record do not score for P)

(1) If the record has 12-15 R,

(*a*) Less than 4P responses is scored 2.

(*b*) Less than 2P responses is scored 4.

(2) If record has 16-25 R,

(*a*) Less than 4P responses is scored 4.

(*b*) Less than 2P responses is scored 6.

(3) If the record has 26 or more R,

(*a*) Less than 4P responses is scored 5.

(*b*) Less than 2P responses is scored 8.

(4) If the "animal" P response on card VIII is missed, the score is 3.

(5) If the "bat" P response on card V is missed, the score is 3.

Sex and Anatomy Content

(1) 1-2 sex (includes embryo responses) or overt anal responses is scored 5.

(2) 3 or more sex or anal responses is scored 9.

(3) More than 3 anatomical responses is scored 3 (does not include bone responses, but does include pelvic responses).

Other Content

(1) If the "animal" percentage is higher than 60, the score is 3.

(2) If the "animal" percentage is lower than 30, the score is 4.

(3) If 5 or more botany responses are present, the score is 3.

(4) If 4 or more landscape responses are present, the score is 3.

(5) If 3 or more "food" responses are present, the score is 4.

(6) If less than 2 "human" responses are present, the score is 5.

(7) If the number of part "human" responses (Hd) exceeds by 3 to 1 (and at least one H is present) the number of complete "human" responses (H), the score is 5.

Minus Cards

If a card has at least 3 responses and the minus responses exceed the plus responses (no color responses count except FC minus) there is a score of 3, on each such card.

Sex Confusion

If a subject verbalizes obvious conflict about the sexual classification of a person or object

(*a*) On at least two responses, the score is 4.

(*b*) On three responses, the score is 5.

(*c*) On more than three responses, the score is 8.

Bizarre Logic

(1) A contaminatory response is scored 20.

(2) A definite pathological DW is scored 10.

Preseveration

If the same response is given on three successive cards and the response in each case is a minus of very poor quality, the score is 15.

* * *

The *total maladjustment score* is equal to the *sum of the penalty weights given.*

APPENDIX VI

SCORING FOR RORSCHACH LOOSENESS

The following are the *maladjustment* signs (and corresponding weights) used in determining the "looseness" score.

Poor form quality (i.e., low F+).

Confused sequence in order of giving responses.

Contamination or similar kinds of bizarre logic.

Color Naming (CN).

Very poor form quality responses.

Poor quality sexual or anal responses.

Poor quality movement responses.

1 M minus response is scored 2.

2 M minus responses are scored 4.

3 M minus responses are scored 8.

4 M minus responses are scored 10.

Perseveration of any of the kinds described is scored 15.

Loose Color or Color of Poor Quality (C, CF—, FC—)

(1) A pure color response (C) without any form is equal to a score of—

(*a*) 3 on card X.

(*b*) 2 on cards II or III.

(*c*) 2 on card VIII.

(*d*) 1 on card IX.

(2) A color response with some minimum form but of poor quality (CF—) and containing either blood content, fire content or anatomy content gives—

(*a*) A score of 3 on cards II, III, or IX.

(*b*) A score of 4 on cards VIII or X.

(3) Minus FC responses

(*a*) One FC minus response equals a score of 2.

(*b*) Two FC minus responses equal a score of 6.

(*c*) Three or more FC minus responses equal a score of 8.

Absence of Color

(1) If color response involving a good integration of form and color (FC+)

(*a*) Is absent on cards II and III, the score is 6.

(*b*) Is absent on cards VIII, IX, and X, the score is 9 (if FC given as *last* response on X, the score is 4).

(2) No color in the total record gives the score of 4.

(3) If there is only one FC in the total record and the rest of the color is C and CF minus, the score is 3.

Color Naming (CN)

(*a*) CN on one card equals a score of 12.

(*b*) CN on two or more cards equals a score of 16.

Form Quality (F+)

(1) If F+ is from 57 to 64 inclusive—

(*a*) An R of 15 through 20 (if less than 15 R, at least 9 of responses should be in the F category) will give a score of 5 for an F% 45-55; and a score of 7 for an F% of 56 and over.

(*b*) An R over 20 will give a score of
4 for an F% of 35-45.
6 for an F% of 46-55.
8 for an F% of 56 and over.

(2) If F+ from 50 through 56—

(*a*) An R of 13 through 20 will give a score of
7 for an F% of 35-40.
10 for an F% of 41-50.
12 for an F% of 56 and over.

(*b*) An R over 20 will give a score of
10 for an F% of 35-40.
15 for an F% of 41-50.
20 for an F% of 56 and over.

(3) If F+ is under 50—

(*a*) An R of 10 through 13 will give a score of
8 for an F% of 35-40.
10 for an F% of 41-50.
15 for an F% of 51 and over.

(*b*) An R of 14 through 20 will give a score of
12 for an F% of 35-40.
18 for an F% of 41-50.
25 for an F% of 51 and over.

(*c*) An R over 20 will give a score of
18 for an F% of 35-40.
25 for an F% of 41-50.
30 for an F% of 51 and over.

Very Poor Form Quality

If a response is F minus, is a severe departure in form from the card area used, and contains very peculiar or fantastic content (e.g., "people catching blood in a cup," "dirty uterus," "inside of a womb"), that response is scored 5.

Number of Responses

(1) If there are less than 20 responses, the score is 2.

(2) If there are less than 15 responses, the score is 5.

(3) If there are less than 12 responses, the score is 9.

Average Time of Initial Responses (T/1R)

(1) If the T/1R is 25 through 35 seconds, the score is 2.

(2) If the T/1R is 36 seconds or greater, the score is 3.

Rejection of Cards

(1) Rejection of card I or IV or VI or IX equals a score of 7.

(2) Rejection of card II or III or V or VII or VIII or X equals a score of 9.

(3) Rejection of two cards in sequence results in an extra penalty of 5. (No more than two rejections, one for each of above categories, is counted).

Shading Responses

(1) 1 to 2 pure shading responses (Y) is scored 2.

(2) 3 to 4 pure Y responses is scored 4.

(3) 5 or more pure Y responses is scored 6.

(4) 2 to 3 shading responses with little form (YF) is scored 2.

(5) 4 or more YF responses is scored 5.

White Space Responses (S)

(1) 4 to 5 S responses is scored 2.

(2) 6 or more S responses is scored 4.

Popular Responses (P)

(If less than 12 R in record do not score for P)

(1) If the record has 12-15 R,

(*a*) Less than 4P responses is scored 2.

(*b*) Less than 2P responses is scored 4.

(2) If record has 16-25 R,

(*a*) Less than 4P responses is scored 4.

(*b*) Less than 2P responses is scored 6.

(3) If the record has 26 or more R,

(*a*) Less than 4P responses is scored 5.

(*b*) Less than 2P responses is scored 8.

(4) If the "animal" P response on card VIII is missed, the score is 3.

(5) If the "bat" P response on card V is missed, the score is 3.

Sex and Anatomy Content

(1) 1-2 sex (includes embryo responses) or overt anal responses is scored 5.

(2) 3 or more sex or anal responses is scored 9.

(3) More than 3 anatomical responses is scored 3 (does not include bone responses, but does include pelvic responses).

Other Content

(1) If the "animal" percentage is higher than 60, the score is 3.

(2) If the "animal" percentage is lower than 30, the score is 4.

(3) If 5 or more botany responses are present, the score is 3.

(4) If 4 or more landscape responses are present, the score is 3.

(5) If 3 or more "food" responses are present, the score is 4.

(6) If less than 2 "human" responses are present, the score is 5.

(7) If the number of part "human" responses (Hd) exceeds by 3 to 1 (and at least one H is present) the number of complete "human" responses (H), the score is 5.

Minus Cards

If a card has at least 3 responses and the minus responses exceed the plus responses (no color responses count except FC minus) there is a score of 3, on each such card.

Sex Confusion

If a subject verbalizes obvious conflict about the sexual classification of a person or object

(*a*) On at least two responses, the score is 4.

(*b*) On three responses, the score is 5.

(*c*) On more than three responses, the score is 8.

Bizarre Logic

(1) A contaminatory response is scored 20.

(2) A definite pathological DW is scored 10.

Preseveration

If the same response is given on three successive cards and the response in each case is a minus of very poor quality, the score is 15.

* * *

The *total maladjustment score* is equal to the *sum of the penalty weights given.*

APPENDIX VI

SCORING FOR RORSCHACH LOOSENESS

The following are the *maladjustment* signs (and corresponding weights) used in determining the "looseness" score.

Poor form quality (i.e., low F+).

Confused sequence in order of giving responses.

Contamination or similar kinds of bizarre logic.

Color Naming (CN).

Very poor form quality responses.

Poor quality sexual or anal responses.

Poor quality movement responses.

1 M minus response is scored 2.

2 M minus responses are scored 4.

3 M minus responses are scored 8.

4 M minus responses are scored 10.

Perseveration of any of the kinds described is scored 15.

BIBLIOGRAPHY

1. ALLPORT, G. W. *Personality.* New York: Henry Holt, 1937.
2. ANGYAL, A. F. Speed and pattern of perception in schizophrenic and normal persons, *Character and Pers.,* 1942, 11, 108-127.
3. ANGYAL, A. F. The diagnosis of neurotic traits by means of a new perceptual test, J. of Psychol., 1948, 25, 105-135.
4. BECK, S. J. *Rorschach test. I. Basic processes.* New York: Grune and Stratton, 1944.
5. BENNETT, G. Structural factors related to the substitute value of activities in normal and schizophrenic persons: I. A technique for the investigation of central areas of personality. II. An investigation of central areas of the personality, *Character and Pers.,* 1941, 10, 42-50. 1942. 10, 227-245.
6. CATTELL, R. B. On the measurement of perseveration, *Brit. J. Educ. Psychol.,* 1935, 5, 76-92.
7. CATTELL, R. B. Perseveration and personality: Some experiments and a hypothesis, *J. Ment. Sci.,* 1935, 81, 151-167.
8. DOWNEY, J. *The will-temperament and its testing.* New York: World Book Company, 1923.
9. DUNLAP, K. Improved forms of steadiness tester and tapping plate, *J. Exp. Psychol.,* 1922, 4, 430-433.
10. FREUD, S. *A general introduction to psychoanalysis.* New York: Garden City Publishing Co., Inc., 1943.
11. GOLDSTEIN, K. Concerning rigidity, *Character and Pers.,* 1943, 11, 209-226.
12. GUILFORD, J. P. *An inventory of factors STDCR (Manual of directions and norms).* Beverly Hills, California: Sheridan Supply Co., 1940.
13. KOUNIN, K. Experimental studies of rigidity, *Character and Person,* 1941, 9, 251-282.
14. LEWIN, K. *A dynamic theory of personality.* New York: McGraw-Hill Book Co., 1935.
15. PINARD, J. W. Tests of preseveration: I, Their relation to character, *Brit. J. Psychol.,* 1932, 23, 5-19.
16. RAPPAPORT, D. *Diagnostic psychological testing,* Vol. 1. New York: New Book Publishers, Inc., 1945.
17. RICKERS-OVSIANKINA, M. Studies of the personality structure of schizophrenic individuals: I. The accessibility of schizophrenics to environmental influences. II. Reaction to interrupted tasks, *J. Gen. Psychol.,* 1937, 16, 153-178, 179-196.
18. RORSCHACH, H. *Psychodiagnostiks.* New York: Grune and Stratton, Inc., 1942.
19. ROSENZWEIG, S. The picture association method and its application in a study of reaction to frustration. *J. Person.,* 1945, 14, 3-23.
20. SHAKOW, D. Schizophrenic and normal profiles of response to an auditory apperceptive test, *Psychol. Bull.,* 1938, 35, 647-652.
21. SHAKOW, D., and ROSENZWEIG, S. I. Play technique in schizophrenia and other psychoses. II. An experimental study of schizophrenic constructions with play material, *Amer. J. Ortho-Psychiat.,* 1937, 7, 36-47.
22. SPEARMAN, C. *The abilities of man: Their nature and measurement.* London: The Macmillan Company, 1927.
23. STEPHENSON, W. An introduction to so-called motor perseveration tests, *Brit. J. Ed. Psychol.,* 1934, 4, 188-208.
24. SZONDI, *Experimentelle Triebdiagnostik.* Bern: Verlag Hans Huber, 1947.
25. TERMAN, L. and MILES, C. *Manual of information and directions for use of attitude-interest analysis test.* New York: McGraw-Hill Book Company, Inc., 1938.
26. WERNER, H. Subnormal and abnormal forms of rigidity, *J. Ab. & Soc. Psychol.,* 1946, 41, 15-24.
27. WERNER, H. The concept of rigidity: A critical evaluation, *Psychol. Rev.,* 1946, 53, 43-52.

VOLUME 64
NUMBER 2

WHOLE NO. 308
1950

Psychological Monographs:
General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

The Value of an Oral Reading Test for Diagnosis of the Reading Difficulties of College Freshmen of Low Academic Performance

By
CHARLES A. WELLS
American International College
Springfield, Massachusetts

Accepted for publication, July 16, 1949

Price $1.00

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION
1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

COPYRIGHT, 1950, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION

TABLE OF CONTENTS

I. The Nature and Basic Limitations of This Study 1
  A. The Nature of the Study 1
    1. Studies of the Frequency of Oral Errors 1
    2. The Relationship between Reading Comprehension and Knowledge of Sentence Elements as Indicated by Oral Phrasing 2
    3. The Relationship Between Oral Errors and Vocabulary 2
  B. Basic Limitations of the Study 2
  C. Summary 2

II. The Study Group and Instruments of Investigation 4
  A. The Study Group 4
  B. The Development of an Oral Instrument for Investigation 5
    1. The Necessity for a New Oral Reading Test 5
    2. Construction of the Oral Reading Test 5
  C. Tests of Silent Reading Skills 7
  D. Intercorrelations of the Silent Reading and Intelligence Tests 8

III. Presentation of the Findings 9
  A. Guiding Considerations 9
  B. A Study of the Oral Error Scores 9
    1. Omissions and Punctuation Errors 9
    2. Insertions 9
    3. Substitutions 9
    4. Mispronunciations 10
    5. Oral Comprehension 10
    6. Time of Oral Reading 11
    7. Total Oral Errors 11

IV. Evaluation of Findings and Summary 13
  A. Evidence of the Diagnostic Value of Specific Types of Oral Reading Error 13

B. The Relationship of Oral Phrasing and Reading Comprehension ... 13
C. The Relationship of Mental Ability to the Oral and Silent Variables 14
D. The Relationship of Oral Reading to Mastery of Sentence Elements . 14
E. The Relationship of Oral Reading with Vocabulary 14
F. The Relationship of Oral Reading Comprehension to Silent Reading Comprehension 14
G. The Relationship of Time of Oral Reading to Time of Silent Reading 15
H. The Relationship of Total Oral Errors to the Language Tests 15
I. Summary and Conclusions 16

APPENDIX A. FRESHMAN ORAL READING TEST, FORMS A AND B 17

APPENDIX B. ACKNOWLEDGEMENTS TO PUBLISHERS 34

REFERENCES 35

CHAPTER I

THE NATURE AND LIMITATIONS OF THIS STUDY

A. THE NATURE OF THE STUDY

THE source of this investigation lies in the need in educational psychology for exactitude of concept formation. The need in the case of this present research arises from what seems to be a current trend—a trend illustrated by a recent text, which after presenting certain generalizations concerning the oral-silent reading relationship, states that "the oral reading examination is indispensable to a thorough exploration of a student's reading abilities" (8, p. 238). The research interest of such a statement lies in the fact that these writers do not offer experimental findings to establish their contention, nor is supporting evidence at the academic levels for which they write (high school and college) to be found in the literature. In fact, the only justification for the statement which this writer could discover is a conclusion by Swanson (who did not use such a test in his study) that an oral reading test "such as Gray's Oral Reading Paragraphs" should be useful at the college level (15, p. 59). A derived conclusion such as this latter would seem hardly justification for the twenty-seven pages which the text quoted devotes to an analysis of various oral reading errors and their implications for reading-diagnosis and for tutoring.

Such generalizations concerning the diagnostic value of oral reading examinations raise a problem of importance to educators. Since oral reading is an observable activity, and thus allows more extensive observation, direction, and measurement than do the more covert silent reading habits, the finding of a high degree of relationship between the two would lend logic to further examination of the assumption that difficulties in the oral medium might be paralleled by similar difficulties in the silent medium, and that improvement in one might result in improvement in the other. Experimental verification of conclusions such as that quoted in the opening paragraph, would provide workers in the field of reading difficulties with a useful instrument for reading diagnosis and for tutoring.

Outlined briefly below are those areas with which this present study will be concerned.

1. *Studies of the Frequency of Oral Errors*

Such studies seem to be based upon the assumption that the student who makes the most frequent errors when reading orally thereby indicates his generally low level of reading proficiency. Opposed to this notion, Marion Monroe (9) and Cassie Payne (11) several years ago asserted that both good and poor readers make all kinds of oral errors, seemingly indiscriminately, when level of reading proficiency is determined by some standard other than the action which is itself being observed—namely, the frequency of the making of oral errors. However, in spite of such objections, studies of error frequency have continued to appear with ever more detailed analyses of errors and of error positions. If oral errors do not show a substantial degree of correlation with other tests of language skills, the value of studies of oral-error frequency will be open to question.

2. *The Relationship between Reading Comprehension and Knowledge of Sentence Elements as Indicated by Oral Phrasing*

Notions of such a relationship run like a thread through the literature. For example, one finds Sangren stating some years ago that "at times the teacher will ask the child to read orally and will discover that he has no recognition of the sentence as a unit of thought . . ." (13, p. 100). Gray tells us that "in addition the examiner is often impressed with many other defects in oral reading . . . and lack of proper phrasing" (6, p. 323). So also we have Thorndike emphasizing the "weighing of each of many elements in a sentence" (16, p. 323). And recently, Vera Paul has stated that it is lack of proper phrasing which keeps many children from getting meaning from the sentence (10, p. 219). Such statements are of interest in view of the fact that this writer has been unable to discover any published research findings to support them.

3. *The Relationship between Oral Errors and Vocabulary*

This relationship has not yet been investigated at the college level. Because of the substantial correlations found between vocabulary and intelligence test scores (.58 in this study), the finding of a high correlation between oral reading errors and vocabulary, and between oral reading errors and intelligence test scores would suggest certain interesting possibilities.

B. BASIC LIMITATIONS OF THE STUDY

A work of this nature has certain limitations which, while they are intrinsic to studies in this field, are sufficiently often ignored to be deserving of specific mention.

Of basic importance is the disagreement concerning the nature and exact definition of the process which is central to our purpose—the reading process. The particular viewpoint to which one adheres would seem to be influential in determining the type of reading test which one will construct, and in determining the interpretations which will be placed upon the results of that test. The manner in which one's viewpoints may affect test construction and interpretation is illustrated aptly by a statement of Courtis, who some years ago argued that McCall's reading tests were in reality tests of thinking and not of reading. "My thesis is that since there are many types of reading situations, and it is necessary to restrict the term reading to one of them, the term should be used to connote the simplest form of the activity, namely, observational reading. In other words, reading to me means the degree of control possessed by the child over the reading mechanism, or to the excellence of the mechanism itself" (2, p. 291). There is no indication in the literature that Courtis has changed his position since that time.

Holding to an opposed view is the definition of reading in a recent text: "To summarize, reading is a psychological process in which the reader obtains meaning from printed words. . . . But clear images and the recognition of words are not enough; he must determine their meaning in a particular context. Interpretation of meaning involves memory of previous experiences, reasoning, and the reader's purposes and attitudes toward the situation as a whole" (8, p. 45).

It may be gathered from the foregoing that the frame of reference within which an investigator in this area will operate is not a matter of common agreement, and hence, constitutes a basic limitation of such studies. It might be appropriate to add that practically every form of behavior in this study—intelligence, spelling, vocabulary, reading rate, and reading comprehension—is a matter of considerable controversy, and thus, exhibits similar limitations.

C. SUMMARY

This investigation, then, is based upon these specific considerations:

1. There is evident a current tendency to generalize broadly concerning the values of oral reading tests for the diag-

nosis of silent reading difficulties.

2. The problem is created by the fact that there is little experimentation to justify such generalizations for the college level.

3. The goal of this present study is to furnish information in the light of which concepts of the diagnostic value of the oral reading test may be evaluated.

CHAPTER II

THE STUDY GROUP AND INSTRUMENTS OF INVESTIGATION

A. THE STUDY GROUP

THE material for this study has been gathered from the scores of the eighty students whose grades placed them in the lowest academic quarter of a class of 312 freshmen at American International College for the second semester ending in June, 1947. In all, 238 of these 312 were completely tested.

As may be noted from Table 1, range and variance in age and length of time between last school attendance and college entrance are considerable. Until clarification of the subject is available, the influence of these factors upon reading proficiency is largely conjectural.

TABLE 1

AGE, LENGTH OF TIME BETWEEN LAST SCHOOL ATTENDANCE AND COLLEGE, AND INTELLIGENCE

	N	Mean	S. D.	Range
Age (in Months)	80	275.95	59.67	215–420
Time between Last School Attendance and College Entrance (in Months)	80	34.64	26.25	3–147
American Council on Education Psychological Examination, 1946 Edition (Raw Scores)	80	95.39	22.83	53–163

To summarize other descriptive data, the group of eighty students upon whose performance this study is based consisted of 74 males and of six females. Of the men all but fourteen were veterans. Of the women two were veterans. Seventy-five of the eighty had been admitted to the college by certification, two had been admitted by entrance examination, and three had been admitted on the basis of the United States Armed Force Institute Examination.

Data on the language backgrounds of the sample are presented in Table 2. Since participation in the study was not voluntary on the students' part, and since the students as a group, in activities quite removed from this study, evidenced the emotionality indicative of feelings of personal insecurity, it was deemed inadvisable to inquire into data of such a nature as would allow the classification of the students into one of the schema of socio-economic status.

Of interest, also, is the extent of student oral reading, since this might possibly affect accomplishment on a test of oral reading. Of the 303 members of this freshman class, who returned completely filled questionnaires, 260 stated that they read orally *one hour a week or less*. From the forty-three who claimed to read orally from one to fifteen hours a week, inquiry elicited the information

TABLE 2

LANGUAGES SPOKEN IN THE HOMES OF STUDENTS
$N=80$

Language	*N*
Armenian	1
English Only	41
French	5
Greek	1
Italian	14
Polish	10
Yiddish	8

that several were members of a radio work-shop, and that many of the others were engaged in various forms of community group work which necessitated considerable oral reading.

B. The Development of an Oral Instrument for Investigation

1. *The Necessity for a New Oral Reading Test*

For the purposes of a study of oral reading at the college freshman level, there are available only two oral reading tests—a sub-test to be found in Strang's *Examiner's Diagnostic Reading Record* (14) and a sub-test of the *Diagnostic Reading Tests,* prepared by a committee under the chairmanship of Triggs (18). Because of the grade range for which these tests were developed, and because of their limited range of material at the college level, neither of them was considered suitable for use in the present study. Determination of the diagnostic usefulness of a test of oral reading for college freshmen would seem to demand the construction of an oral reading instrument for that purpose.

2. *Construction of the Oral Reading Test*

Since reading disability may be specific to certain types of material, the content of the reading passages should be functional, i.e., should sample the fields of reading in which a college freshman engages in his academic work. For the Freshman Oral Reading Test[1] the fields of reading which were chosen, basing the choice upon the program of general studies for the freshman year at American International College, were psychology, history, the natural sciences, and English. The reading passages were selected from the following sources: from freshman texts which are no longer published because of text revisions, from the Sunday edition of the *New York Times,* and from sophomore texts.

Comprehension questions were devised for the reading passages, and these questions together with the selections were subjected to expert criticism and to revision. In its revised form the test was administered to a preliminary group of eighty students. An item analysis of the comprehension questions was based upon the performance of this preliminary group. Techniques used for the item analysis were those suggested by Davis (3, pp. 6-7), and the difficulty and discrimination indices for the comprehension questions retained for the experimental form of the test are to be found in Tables 3 and 4.

The test was then administered to the eighty members of the study group. Reliabilities between Forms A and B, obtained from this administration, were: for time of oral reading, .90; and for oral comprehension, .67.

Based upon experience with the preliminary group, the following oral reading errors were recorded on the experimental administration of the test:

a. Mispronunciations—in cases in which the mispronunciation was not clearly owing to substitutions, omissions, or insertions within the word.

b. Punctuation Errors—in cases in which change of voice did not indicate heed for the punctuation mark. It was felt that punctuation errors would constitute a useable device for approximating an objective measure of oral phrasing and rhythm.

c. Omissions—in cases of omission both of words and of word parts.

---

[1] The oral instrument devised for this study will be referred to hereafter by this title.

TABLE 3

PERCENTAGE OF SUCCESSES IN THE UPPER 27% AND LOWEST 27% OF THE PRELIMINARY GROUP*

$N=80$

Form A

I-1	I-2	I-3	I-4	II-1	II-2	II-3	III-1
70	34	78	78	68	50	49	55
8	5	41	44	21	11	6	15
III-2	III-3	IV-1	IV-2	IV-3	V-1	V-2	VI-1
64	97	67	61	64	83	83	73
5	58	17	28	18	28	48	25
VI-2	VII-1	VII-2	VII-3	VII-4	VIII-1	VIII-2	VIII-3
54	51	31	43	83	83	97	43
7	20	1	7	31	35	68	2

Form B

I-1	I-2	I-3	II-1	II-2	II-3	III-1	III-2	III-3
52	68	93	68	70	77	60	86	36
20	30	31	25	37	34	16	55	1
III-4	III-5	IV-1	IV-2	V-1	V-2	V-3	VI-1	VI-2
46	74	86	60	48	84	80	38	72
16	43	22	22	11	54	11	12	28
VI-3	VII-1	VII-2	VIII-1	VIII-2	VIII-3	VIII-4	VIII-5	
87	67	86	52	38	76	76	52	
38	4	43	20	12	25	34	17	

* Of each pair of numbers the upper gives the percentage of successes for the upper 27% of the group and the lower the percentage for the lowest 27%.

d. Substitutions—in cases of substitutions of both words and of word parts.

e. Repetitions—including any number of repetitions of a word. Even a single repetition may be indicative of oral difficulties.

f. Insertions—in cases of insertions of both words and word parts.

g. Hesitations—A liberal allowance of three seconds was placed upon the allowable limits of a hesitation in order to allow for variations in breath control, attention span, and so on.

h. Word Stress—While it is by no means clear that this error would be of extended significance in a consideration of reading disability, since its probable origin may lie in the world's being more a part of the reader's reading vocabulary, than it is of his speaking vocabulary, the practice occurred a sufficient number of times to justify its inclusion as an error to be scored on the experimental form of the test.

As a device for choosing the oral errors which were to be used in the research, scatter diagrams of oral error performance were plotted. The essential material from these scattergrams is presented in Table 5.

It was decided arbitrarily that if over fifteen percent of the sample obtained a zero error score on any specific error, this error would not be suitable for inclusion in the study. As may be noted from Table 5, Mispronunciations, Substitutions, and Omissions warrant inclusion. Word Stress errors, Hesitations, and Repetitions do not constitute errors useful for our purposes. Insertions and Punctuation Errors are of questionable value, but will be included.

TABLE 4

DAVIS'S DISCRIMINATION AND DIFFICULTY INDICES FOR THE COMPREHENSION QUESTIONS OF THE FRESHMAN ORAL READING TEST*

$N=80$

Form A							
I-1	I-2	I-3	I-4	II-1	II-2	II-3	III-1
47	30	24	23	31	30	38	30
44	32	55	56	47	40	38	42
III-2	III-3	IV-1	IV-2	IV-3	V-1	V-2	VI-1
49	42	35	21	31	38	24	33
42	66	46	47	45	53	58	49
VI-2	VII-1	VII-2	VII-3	VII-4	VIII-1	VIII-2	VIII-3
39	22	45	36	35	33	35	45
40	42	36	36	54	54	69	34

Form B								
I-1	I-2	I-3	II-1	II-2	II-3	III-1	III-2	III-3
22	24	48	28	21	28	31	24	48
42	49	56	48	52	53	44	61	31
III-4	III-5	IV-1	IV-2	V-1	V-2	V-3	VI-1	VI-1
22	21	45	25	30	22	50	21	22
40	54	52	43	38	60	48	36	53
VI-3	VII-1	VII-2	VIII-1	VIII-2	VIII-3	VIII-4	VIII-5	
35	53	30	22	21	34	28	25	
56	42	58	42	36	51	33	42	

* Of each pair of numbers the discrimination index is the upper, the difficulty index is the lower.

C. TESTS OF SILENT READING SKILLS

The guiding considerations in the choice of tests of silent reading and of intelligence with which to correlate the Freshman Oral Reading Test were: (1) that they should be of college level, (2) that there should be information in the literature as to their reliabilities and usefulness, and (3) that their administration and scoring should not be so complex as to render their use impracticable under the conditions in which this investigation was conducted.

In the light of these considerations

TABLE 5

ERROR PERFORMANCE OF THE SAMPLE ON FORMS A AND B OF THE ORAL READING TEST

$N=80$

Type of Oral Error	Mean Forms A and B	S. D. Forms A and B	Mdn. Forms A and B	r between Forms A and B
Word Stress	.8	1.1	.9	.11
Hesitation	1.3	1.7	.9	.30
Punctuation Errors	3.5	4.2	2.2	.77
Insertions	1.9	1.5	2.5	.29
Substitutions	3.0	2.6	3.0	.59
Omissions	3.8	2.8	3.7	.35
Mispronunciations	5.2	4.6	4.5	.86
Repetitions	3.6	4.5	2.4	.70

the following examinations were selected.

1. *Intelligence*—American Council on Education Psychological Examination for College Freshmen, 1946 Edition.

2. *Vocabulary*—Cooperative English Test C 2: Reading Comprehension (Higher Level), Form Q, Part 1.

3. *Rate and Level of Comprehension*—Cooperative English Test C 2: Reading Comprehension (Higher Level), Form Q, Part II.

4. *Speed of Reading*—Minnesota Speed of Reading Test for College Students, Form B.

5. *Sentence Structure*—The Purdue Placement Test in English for High Schools and Colleges, Form A, Part IV.

6. *Grammatical Classification*—The Purdue Placement Test in English for High Schools and Colleges, Form A, Part II.

7. *Recognition of Grammatical Errors*—The Purdue Placement Test in English for High Schools and Colleges, Part III.

D. INTERCORRELATIONS OF THE SILENT READING AND INTELLIGENCE TESTS

In order to determine whether or not these tests of language skills and of intelligence could be presumed to measure relatively independent abilities, performance on them by the students for whom complete results were obtained was intercorrelated. Table 6 presents the results of such intercorrelation. All correlations are of raw scores.

The list preceding Table 6 identifies the numbers used for coding the language tests included. The means and the standard deviations presented at the foot of each column are these measures for the scores obtained on that test the code number for which stands at the head of the column.

TABLE 6

INTERCORRELATIONS OF THE CRITERION TESTS

$N = 238$

1. Cooperative English Test C 2, Vocabulary Score
2. Cooperative English Test C 2, Speed of Comprehension
3. Cooperative English Test C 2, Level of Comprehension Score
4. Cooperative English Test C 2, Total Score
5. Minnesota Speed of Reading, Form B
6. Purdue Placement, Grammatical Classification Score
7. Purdue Placement, Recognition of Grammatical Errors Score
8. Purdue Placement, Sentence Structure Score
9. A.C.P.E., "Q" Score
10. A.C.P.E., "L" Score
11. A.C.P.E., Total Score

	1	2	3	4	5	6	7	8	9	10	11
1		.844	.596	.840	.432	.385	.266	.274	.239	.692	.580
2			.818	.985	.542	.212	.315	.383	.396	.712	.675
3				.852	.455	.355	.240	.328	.394	.583	.629
4					.499	.393	.296	.353	.403	.678	.661
5						.143	.251	.264	.339	.472	.490
6							.494	.366	.199	.811	.386
7								.238	.135	.344	.321
8									.129	.300	.354
9										.524	.817
10											.874
11											
M.	20.9	17.4	11.6	49.9	14.1	11.4	19.5	16.7	41.4	68.0	108.9
S.D.	9.9	10.6	7.4	25.5	5.2	5.1	9.2	4.8	11.3	8.5	23.1

Chapter III

PRESENTATION OF THE FINDINGS

A. Guiding Considerations

The critical ratio is used as a test of the significance of the r's presented in this section. The convention which has been followed is that ratios between 1.96 and 2.58 may be regarded as being significant, and that those ratios which are larger than 2.58 are to be regarded as being very significant. Kelley's "one-third sigma rule" has been followed in reporting the findings of the study (7, p. 223).

B. A Study of the Oral Error Scores

The following Tables, 7 to 14, present the results obtained from correlating the oral errors—Mispronunciations, Substitutions, Omissions, Insertions, and Punctuation Errors, as well as scores of oral comprehension, total oral errors, and time of oral reading—with the silent reading measures used in the investigation. These oral test scores are based upon the use of both Forms A and B of the Freshman Oral Reading Test considered as if they together constituted but one examination.

1. *Omissions and Punctuation Errors*

As is to be noted from an inspection of Tables 7 and 8, neither the omission of words or word parts nor the extent of disregard for punctuation marks during oral reading are related significantly to any of the silent tests of language skills used in this study. This is a finding of interest since the Punctuation Error score was intended as an objective measure of phrasing during oral reading.

TABLE 7

Correlations of Omissions Errors with Other Test Results

Test	r*
Cooperative Speed of Comprehension	.09
Cooperative Level of Comprehension	.16
Cooperative Vocabulary	.17
Cooperative Total	.14
Minnesota Speed of Reading	.11
Purdue Sentence Structure	.08
Purdue Recog. Gram. Errors	.15
Purdue Gram. Class.	.10
American Council Psychological	.07
Total Oral Comprehension	.03

* None of the r's are significant at the 5% level.

TABLE 8

Correlations of Punctuation Errors with Other Test Results

Test	r*
Cooperative Speed of Comprehension	.09
Cooperative Level of Comprehension	.04
Cooperative Vocabulary	.02
Cooperative Total	.03
Minnesota Speed of Reading	.11
Purdue Sentence Structure	.11
Purdue Recog. Gram. Errors	.03
Purdue Gram. Class.	−.04
American Council Psychological	.07
Total Oral Comprehension	−.15

* None of the r's are significant at the 5% level.

2. *Insertions*

The insertion of added sounds and words during oral reading correlates significantly with the results of but one test, the Cooperative Reading Vocabulary Test (Table 9).

3. *Substitutions*

The relationships obtained between this type of error—the substitution of words or word parts during oral reading—and the silent tests were of interest to this writer in view of the importance given to this error in several of the studies of the value of oral reading. However, as is evident from Table 10, Substitutions are correlated significantly with none of the criterion scores except with those of the Purdue tests of Gram-

TABLE 9

CORRELATIONS OF INSERTIONS ERRORS WITH OTHER TEST RESULTS

Test	r	σ_r
Cooperative Speed of Comprehension	.11	.11
Cooperative Level of Comprehension	.19	.11
Cooperative Vocabulary	.27	.10*
Cooperative Total	.10	.11
Minnesota Speed of Reading	.18	.11
Purdue Sentence Structure	.12	.11
Purdue Recog. Gram. Errors	.13	.11
Purdue Gram. Class.	.17	.11
American Council Psychological	.06	.11
Total Oral Comprehension	.05	.11

* $P < .01$.

TABLE 10

CORRELATIONS OF SUBSTITUTIONS ERRORS WITH OTHER TEST RESULTS

Test	r	σ_r
Cooperative Speed of Comprehension	.01	.11
Cooperative Level of Comprehension	.15	.11
Cooperative Vocabulary	.10	.11
Cooperative Total	.08	.11
Minnesota Speed of Reading	.12	.11
Purdue Sentence Structure	−.06	.11
Purdue Recog. Gram. Errors	.21	.11
Purdue Gram. Class.	.22	.11*
American Council Psychological	.05	.11
Total Oral Comprehension	−.07	.11

* $.05 > P > .01$.

matical Classification and Recognition of Grammatical Errors.

4. *Mispronunciations*

Of the errors studied in this investigation, Mispronunciations (Table 11), consisting of gross errors in the pronunciation of words, are correlated with more of the tests of silent reading skills than are any of the other oral error scores used in the study. Such errors show a significant correlation with the tests of the level of silent reading comprehension, vocabulary, and the psychological examination. To be noted is the fact that there is a significant difference in the size of the correlation between this error and vocabulary compared to the size of the correlation between Mispronunciations and level of silent comprehension (12, p. 185). With regard to the latter the correlations between Mispronunciations and level of silent comprehension and mispronunciations and oral comprehension show no significant difference.[1]

TABLE 11

CORRELATIONS OF MISPRONUNCIATIONS ERRORS WITH OTHER TEST RESULTS

Test	r	σ_r
Cooperative Speed of Comprehension	.21	.11
Cooperative Level of Comprehension	.29	.10†
Cooperative Vocabulary	.42	.09†
Cooperative Total	.38	.10†
Minnesota Speed of Reading	.12	.11
Purdue Sentence Structure	.03	.11
Purdue Recog. Gram. Errors	.10	.11
Purdue Gram. Class.	.18	.11
American Council Psychological	.29	.10†
Total Oral Comprehension	.25	.11*

* $.05 > P > .01$.
† $P < .01$.

In view of the size of the correlation between Mispronunciations and the psychological examination, the problem of the possible role played by mental ability in the relationship observed between Mispronunciations and the other variables is relevant. Holding intelligence constant (12, p. 243), significant partial *r's* are obtained between Mispronunciations and vocabulary, and between Mispronunciations and oral comprehension.[2]

5. *Oral Comprehension*

Oral comprehension, as measured by

---

[1] $r_{12} - r_{13} = .04$; $\sigma r_{12} - r_{13} = .12$

[2] Mispronunciations and Level of Silent Comprehension:

$$r_{01.2} = .14;\ \sigma r_{01.2} = .11$$

Mispronunciations and Vocabulary:

$$r_{01.2} = .22;\ \sigma r_{01.2} = .10$$

Mispronunciations and Oral Comprehension:

$$r_{01.2} = .18;\ \sigma r_{01.2} = .08$$

TABLE 12

CORRELATIONS OF ORAL COMPREHENSION WITH OTHER TEST RESULTS

Test	r	σ_r
Cooperative Speed of Comprehension	.42	.09†
Cooperative Level of Comprehension	.39	.10†
Cooperative Vocabulary	.38	.10†
Cooperative Total	.45	.09†
Minnesota Speed of Reading	.08	.11
Purdue Sentence Structure	.17	.11
Purdue Recog. Gram. Errors	.18	.11
Purdue Gram. Class.	.13	.11
American Council Psychological	.25	.11*

* $.05 > P > .01$.
† $P < .01$.

the correctly answered check items of the Freshman Oral Reading Test, is related significantly with the three part scores and the total score of the Cooperative English Reading Comprehension Test and with the psychological examination (Table 12). With mental ability held constant, the resulting partial r's of these variables are significant.[3]

6. *Time of Oral Reading*

The results of the correlation of this oral score with the other measures used in the study are presented in Table 13. Of interest is the observation that there is not a significant difference between the correlations of Time of Oral Reading and time of silent reading as measured, on the one hand, by the Cooperative Speed of Comprehension score and, on the other hand, by the Minnesota Speed of Reading results.[4]

TABLE 13

CORRELATIONS OF TOTAL ORAL TIME OF READING WITH OTHER TEST RESULTS

Test	r	σ_r
Cooperative Speed of Comprehension	.38	.10†
Cooperative Level of Comprehension	.19	.11
Cooperative Vocabulary	.25	.11*
Cooperative Total	.31	.10†
Minnesota Speed of Reading	.33	.10†
Purdue Sentence Structure	.03	.11
Purdue Recog. Gram. Errors	.07	.11
Purdue Gram. Class.	.06	.11
American Council Psychological	.10	.11
Oral Comprehension	.00	.00

* $.05 > P > .01$.
† $P < .01$.

7. *Total Oral Errors*

As with Substitutions so with Total Oral Errors we have a score which is of interest in view of the fact that its relationships have constituted one of the major findings of several related studies. For example, Gilmore, who used an oral test of the same general type as that of this investigation, with a group consisting of practically all of the students in the first eight grades of one school system, found a mean r of .67 between total errors and silent reading comprehension

TABLE 14

CORRELATIONS OF TOTAL ORAL ERRORS WITH OTHER TEST RESULTS

Test	r	σ_r
Cooperative Speed of Comprehension	.14	.11
Cooperative Level of Comprehension	.27	.11*
Cooperative Vocabulary	.34	.10†
Cooperative Total	.29	.10†
Minnesota Speed of Reading	.18	.11
Purdue Sentence Structure	.08	.11
Purdue Recog. of Gram. Errors	.23	.11*
Purdue Gram. Class.	.15	.11
American Council Psychological	.20	.11
Total Oral Comprehension	.19	.11
Total Oral Time	.38	.10†

* $.05 > P > .01$.
† $P < .01$.

---

[3] Oral Comprehension and Speed of Silent Comprehension:

$r_{01.2} = .35$; $\sigma r_{01.2} = .10$

Oral Comprehension and Level of Silent Comprehension:

$r_{01.2} = .31$; $\sigma r_{01.2} = .10$

Oral Comprehension and Vocabulary:

$r_{01.2} = .30$; $\sigma r_{01.2} = .10$

Oral Comprehension and Cooperative Total Score:

$r_{01.2} = .40$; $\sigma r_{01.2} = .10$

[4] $r_{12} - r_{13} = .05$; $\sigma r_{12} - r_{13} = .10$

for the eight grades (5, p. 77). Swanson, who worked with two groups of college freshmen, one of poor and one of good readers, and who recorded their oral reading of three paragraphs, found an r of .44 between total oral errors and silent reading comprehension (15, p. 52). In contrast, in this present study, which is based upon the use of an oral reading test with eighty college freshmen of low academic achievement, the correlations between Total Oral Errors and both oral comprehension and the psychological examination are not significant at the 5% level of confidence, although they are almost of a size to be acceptable, .20 and .19. Nor is this error score correlated significantly with speed of silent comprehension, nor with the tests of sentence structure, except for the Purdue test of Recognition of Grammatical Errors. Total Oral Errors are related equally with level of silent comprehension and with vocabulary.[5] This score is significantly correlated, as well, with time of oral reading. As may be noted from Table 14 the significant correlations are: with level of comprehension, .27; with vocabulary, .34; with the Cooperative total score, .29; with recognition of grammatical errors, .23; and with time of oral reading, .38.

The size of these correlations when compared to those obtained by Gilmore (Table 15) leads to a consideration of a probable relationship between Total Oral Error Scores and grade level.

According to the tendency evidenced by Gilmore's findings (Table 15), one might well expect to find a correlation between Total Oral Errors and silent reading comprehension which would be lower at the college freshman level than the r of .56 found by Gilmore at the eighth grade. That Gilmore may have suspected the existence of such a trend is suggested by his comparison of his findings with the r of .44 obtained by Swanson between the same variables, and his conclusion that oral reading habits are more closely related to silent reading comprehension at the elementary than at the college level (5, p. 154).

Regressing Gilmore's correlations upon their grade levels, and weighting the coefficients in order to provide for differences in grade size (7, p. 352), yields a regression coefficient for correlation coefficient regressed on grade of $-.0431$. $F_{1(N-2)}$ equals 8.3014 (7, p. 358), t equals 2.88 (7, p. 357), and thus, we have a basis for rejecting the hypothesis that the differences in size to be observed among the correlations obtained by Gilmore are ascribable to chance deviations from a true correlation of zero between coefficient of correlation and grade. A suggested corollary would seem to be the decreasing value of a test of oral reading as a source of diagnostic clues concerning the nature of silent reading comprehension, as grade level advances.

TABLE 15

CORRELATIONS AT GRADES II THROUGH VIII BETWEEN ORAL READING SCORES AND COMPREHENSION SCORES ON THE STANFORD TEST (5, p. 110)*

Grade	Correlation between Oral Reading Errors and Comprehension Scores on the Stanford Test (Part I)
II	918
III	631
IV	693
V	693
VI	659
VII	572
VIII	561

* $N=138$ for Grades II and III, $N=184$ for Grades IV, V, and VI, and $N=138$ for Grades VII and VIII.

---

[5] $r_{12} - r_{13} = .07$; $\sigma r_{12} - r_{13} = .06$

CHAPTER IV

EVALUATION OF FINDINGS AND SUMMARY

A. EVIDENCE OF THE DIAGNOSTIC VALUE OF SPECIFIC TYPES OF READING ERRORS

OF THE oral errors used in this study Omissions and Punctuation Errors are related to none of the silent language tests. Insertions are correlated .27 with vocabulary. Substitutions errors are related with but two of the silent tests—.22 with the Purdue Test of Grammatical Classification, and .22 with the Purdue Recognition of Grammatical Errors. The Mispronunciations error, considered from the viewpoint of its diagnostic usefulness, as determined by the number of tests with which it is correlated, offers the most promise. Its correlation with oral comprehension is .25, with silent comprehension .29, and with vocabulary is .42.

These figures would seem to suggest that gross mispronunciations during oral reading are related to some extent with the vocabulary level of the reader. However, it is difficult to conceive of a manner in which these oral errors might reveal information about a student's vocabulary difficulties which could not be more easily and directly obtained from a vocabulary test itself.

In general, it may be said that in view of the possible number of significant correlations which might have been obtained between specific types of oral error and tests of language skills, the number actually obtained is extremely small. This finding, when taken in conjunction with Gilmore's discovery at the elementary level that of the seven types of oral errors which he studied, only two—Substitutions and Words Pronounced by the Examiner—are significantly related to silent reading comprehension (5, pp. 145-6), would suggest that specific types of oral error are of but indifferent value for the furnishing of clues in the diagnosis of reading difficulties, either oral or silent. This would further suggest that such analyses of specific oral errors and their diagnostic significance as those of McCullough, Strang, and Traxler (8, pp. 246-273) have little sound basis.

B. THE RELATIONSHIP OF ORAL PHRASING AND READING COMPREHENSION

The Punctuation error, consisting of the student's failure to change the voice appropriately, was intended as an objective measure of the student's oral phrasing. In lieu of some such measure it would seem that estimates of the subject's skill in oral phrasing would tend to be somewhat subjective. However, insofar as Punctuation Errors served this purpose, no evidence was furnished by this study to indicate that phrasing during oral reading is related to either silent or oral comprehension. One would not wish to push the suggested conclusion overly far in view of the suspected inadequacy of the Freshman Oral Reading Test as a measure of this error. However, the logic of such a conclusion would be supported by the not infrequent observation of students who, to all appearances, read quite well orally with little or no comprehension of the material read. (In this connection we might cite the correlations of .27 between Total Oral Errors and level of silent reading comprehension, and .19

between Total Oral Errors and Oral Comprehension.)

C. The Relationship of Mental Ability to the Oral and Silent Variables

The influence of mental ability upon the relationships between the oral and silent scores has been determined by holding mental ability constant and testing the significance of the consequent partial *r*'s. In only one case, that of the correlation between Mispronunciations and level of silent reading comprehension, was the resulting partial *r* nonsignificant. This would suggest that although mental ability is a factor in the oral-silent reading relationship, this relationship may be explained in terms other than those of the influence of a factor of mental ability common to the two variables.

D. The Relationship of Oral Reading to Mastery of Sentence Elements

The two instances of relationship between oral errors and tests of mastery of sentence structure have been mentioned previously. These are the correlations of .22 between Substitutions and grammatical classification, and of .22 between Substitutions and recognition of grammatical errors. These *r*'s would seem to indicate an extremely slight tendency among freshmen of low academic achievement who are uncertain about the structural components of an English sentence to betray that uncertainty by the making of substitutions in their oral reading.

Of seemingly greater importance is the failure in this study to find a significant correlation between tests of mastery of sentence elements and oral comprehension. If this relationship were to be found, one implication might be the indicated value of instruction in the elements of sentence structure as a part of a program of instruction in reading. Our findings would not seem to serve to justify this suggestion.

E. The Relationship of Oral Reading with Vocabulary

The correlation between Mispronunciations and vocabulary is .42; between Time of Oral Reading and vocabulary it is .34. Between Oral Comprehension and vocabulary the *r* is .38, which is not significantly different from the *r* of the same oral variable with level of silent reading comprehension. This would suggest that oral reading comprehension is related about equally both with silent reading comprehension and with vocabulary.

F. The Relationship of Oral Reading Comprehension to Silent Reading Comprehension

Oral reading comprehension is, of the oral variables studied, related most substantially to silent reading comprehension. The *r* between Oral Reading Comprehension and the total score of the Cooperative English Reading Comprehension Test C 2 is .45. (This latter score is a composite score representing performance on tests of speed of silent reading comprehension, level of silent reading comprehension and vocabulary.) In view of the possible effects of differing methods of testing reading comprehension upon the *r*'s obtained between two tests, which have been pointed out by Eurich (4, pp. 404-6) and Tinker (17, pp. 217-18), we may be justified in regarding this *r* as being a somewhat too low indication of the true relationship existing between the two variables. Consequently, it would seem safe to suggest that this *r* indicates that college freshmen of low achievement who understand well what they read orally tend also to understand well what they read silently.

It may be remarked, however, that this observation will be hardly surprising to those who work in the field of reading difficulties, and would seem to have little bearing upon the problem of whether or not difficulties in word attack during oral reading are indicative of similar difficulties in word attack during silent reading.

G. The Relationship of Time of Oral Reading to Time of Silent Reading

Time of Oral Reading is correlated significantly with speed of silent comprehension, the *r* being .45. It is not significantly correlated with level of silent comprehension. With Total Oral Errors, Time of Oral Reading is correlated .38, and with vocabulary, .25. The coefficient of .38 between Time of Oral Reading and Total Oral Errors may be compared to the *r* of .56 found by Swanson (15, p. 46). These findings would seem to indicate a degree of relationship between speed in silent reading and speed in oral reading. (However, any consideration of an application of this finding would involve, as well, the observation that Time of Oral Reading is not significantly correlated with Oral Comprehension.)

H. The Relationship of Total Oral Errors to the Language Tests

This error score, which formed the main oral measure of Gilmore's study of a general elementary school sampling, showed lower evidence of relationship at the college freshman level with students of low academic performance. For example, significant correlations were not found between this error and mental ability for the freshman group of this study. The correlation of Total Oral Errors with level of silent reading comprehension in this present study is .27, and between this error and vocabulary is .34. These two correlations are not significantly different in size.

These *r*'s of .27 and .34 are somewhat lower than the *r* of .56 which Gilmore found between total oral errors and silent reading at the eighth grade. However, as has been seen from an examination of Gilmore's *r*'s for these two variables from grades two through eight, there is a steadily decreasing trend in the size of the *r*'s as the grade level advances. The hypothesis has been tested that the differences to be observed in the size of the correlations obtained by Gilmore represent chance differences from a true correlation of zero between correlation coefficients and grade level, and it has been observed that Gilmore's *r*'s are significantly related to their grade levels. It seems reasonable, consequently, to infer that the lower *r*'s found in this present investigation between Total Oral Errors and silent reading comprehension represent correlations of such a size as might be expected in the light of the decreasing size of the *r*'s between total oral errors and silent reading comprehension as the grade level advances.

This observation leads to certain further considerations. The first of these is that since the shift from oral to silent reading is made in the second grade, it would not seem to be illogical to presume that habits of beginning silent reading will be largely those of oral reading. This would serve to account for the *r* of .92 found by Gilmore between the two variables at the second grade. It would further seem to indicate that diffi-

Form A

I

Excessively long study periods are to be avoided on account of the diminishing returns which result from fatigue. But few students over-exert themselves in matters of study, although many commence their daily program in a tired condition. Their fatigue may frequently be attributed to preoccupation with details and to conflicts arising from excessive indulgence in social activities, or from long hours spent in occupational endeavors. Odd moments and periods between classes should be utilized, but a workable plan makes provision for periods long enough for concentrated and sustained effort.

1. In the final analysis, excessively long study periods are to be avoided because of
    a. occupational endeavors
    b. lack of purpose
    c. diminishing returns
    d. tiredness
    e. shortened social activities
2. This writer claims that conflicts arise from
    a. diminishing returns
    b. lack of plan
    c. uncertainty
    d. excesses
    e. personal problems
3. This passage is an argument for
    a. non-indulgence in social activities
    b. planned study
    c. not studying too long
    d. guarding against fatigue
    e. not holding an outside job
4. The chief point discussed in the paragraph is
    a. the results of outside work
    b. the desirability of a workable plan of study
    c. the desirability of hard work in studies
    d. the result of social activities
    e. the result of fatigue

II

Right belief and credulity refer to habits of mind as well as to standards of evidence. "Thinking straight is essential to seeing straight. The evidence grows out of the attitude far more than the attitude results from the evidence." Some people are so entrenched in their own convictions that they resist the truth. It is easy for an anti-vivisectionist, for example, to discredit the values of animal dissection in the progress of medical research. Most people believe too much rather than too little. Credulity is rampant. Ours may be a scientific age, yet this era is the hey-day of the fortune tellers.

1. The writer believes that this is an age of
   a. over belief
   b. anti-vivisectionists
   c. medical research
   d. straight thinking
   e. right belief
2. Which of the following statements is not to be found in the passage?
   a. This may be a scientific age.
   b. People consult fortune tellers.
   c. Thinking straight is necessary for seeing straight.
   d. Most people believe too little.
   e. Anti-vivisectionists discredit animal research.
3. This writer is disturbed about people's mental habits.
   a. Yes
   b. No
   c. Passage does not say

III

We are today in the midst of those convulsions. Germany is divided and broken, slowly starving, a cancerous growth in the heart of Western Europe, a breeding place for the dark philosophy of Nietzche. There festers in her ruins the poison of fascism, the virus of communism. Either totalitarianism flourishes in decadence and destruction. In the Netherlands, the wealth of the Indies no longer builds stately mansions; scores of thousands of Dutch want to emigrate from broken Europe. Belgium is scarred by old wounds and ancient animosities, and France—still politically bankrupt and ethically and intellectually confused—lies weak and torn and divided.

1. The writer of this passage is best described as being
   a. optimistic
   b. dubious
   c. discouraged
   d. pessimistic
   e. excited
2. The author compares Germany to
   a. Nietzche
   b. a political bankrupt
   c. a progressive disease
   d. a stately mansion
   e. a ruin
3. In the ruinous prospect which faces Western Europe the author finds hope for
   a. France
   b. The Netherlands
   c. Belgium
   d. Germany
   e. no country

IV

The first attack on merchantilism developed in France in the eighteenth century. Men known as physiocrats began to argue that governments would do well to leave business alone. Deducing their ideas from the Newtonian faith in natural law, they believed that there were natural laws in economics, which were as immutable as the natural laws of astronomy. Laissez faire (let things alone)—let nature run its course—and all will be well, these men proclaimed. The physiocrats were very theoretical in their beliefs, and living in an agricultural country, they considered agriculture to be the principle source of all wealth.

1. The physiocrats believed that economic forces are the result of
    a. agriculture
    b. leaving business alone
    c. natural laws
    d. theoretical beliefs
    e. governmental regulations
2. The physiocrats opposed
    a. Newton
    b. mercantilism
    c. laissez faire
    d. governments
    e. agriculture
3. The writer finds a parallel to the physiocrat's view of economics in
    a. astronomy
    b. agriculture
    c. physics
    d. biology
    e. mercantilism

VI

Among the earliest cells to become differentiated in an elongating internode are those immediately beneath the epidermis. The cells enlarge somewhat, and later their protoplasts secrete a few additional layers of wall material. Such thickening of the walls results in the formation of a mechanical tissue of considerable rigidity. This tissue doubtless functions in support during the early stages of development. In an elongating internode, at some distance inward from the surface of the stem, strands of cells are differentiated that extend vertically, and parallel with one another through the internode. These strands, composed chiefly of elongated cells, will eventually mature into vascular bundles.

---

1. The process described in this passage is one of
   a. cell composition
   b. decay
   c. thickening
   d. function
   e. development
2. Mechanical tissue is formed by the thickening of the
   a. walls
   b. bundles
   c. meristematic cells
   d. strands
   e. internodes

V

Given adequate training the scientist focuses his attention on some particular problem in his chosen field. He analyzes and studies the problem carefully. He determines the relationship to it of known facts and principles. While he may forecast in his mind a probable result of his work, he does not let his expectations prejudice his method, and he is never surprised if the expectations be not fulfilled. He analyzes the problem into constituent factors, subjects each to observation and to experiment, and records with painstaking accuracy each discovered aspect, phenomenon, or fact. To obviate the chance of error, he repeats.

---

1. Error in experimentation is obviated, according to this passage by
   a. forecasting results
   b. thorough training
   c. repetition
   d. analysis
   e. preliminary study
2. The purpose of this writer is to
   a. warn
   b. describe
   c. persuade
   d. admonish
   e. entertain

VII

One wishes to avoid exaggeration, but it seems safe to say that there are times, and not infrequently, when the spirit of comedy makes life endurable. The tragedies of life, because they are serious, we can bear with fortitude, but they are for the most of us rare, and to some they never come. But the lesser bafflements, the grotesque inconsistencies, the irritating irrelevancies—these are the things that though individually they may be of no power to quell, in the aggregate are like sand in one's shoes, a sliver under the skin, a cinder in the eye. It is comic laughter that wipes the slate clean and refreshes the soul.

1. The writer discusses
    a. fortitude
    b. laughter
    c. inconsistencies
    d. comedy
    e. tragedy
2. Tragedy is
    a. a grotesque experience
    b. a baffling experience
    c. a cumulative experience
    d. a common experience
    e. an infrequent experience
3. The most irritating factors in human experience are to be found in
    a. lack of human fortitude
    b. major tragedies
    c. lesser inconsistencies
    d. grotesque bafflements
    e. minor annoyances
4. Using one or two words, tell what the writer recommends as a method for relieving irritation.

VIII

Probably at no other time in the world's history has the individual been so much occupied with himself as in the years through which we have recently been living. The mark of that self-absorption has been deep upon our literature. The state of mind which it represents has been especially acute among men and women of intellectual pretensions, but the malady has by no means been confined to them. If it had been there would have been nothing like the market which has existed during recent years for the books of self-help which have appeared in such profusion.

1. According to this passage the concern of many individuals today is with
    a. historical trends
    b. intellectual problems
    c. acute states of mind
    d. themselves
    e. politics
2. One indication that this concern is wide spread is the large sale of
    a. books on philosophy
    b. books on self-help
    c. the Bible
    d. books on current history
    e. books on economic theory
3. According to this passage, never before in the history of our nation have we been so concerned with the needs of others.
    a. Yes
    b. No
    c. Passage does not say

Form B

I

Some courses, because of either intrinsic difficulty or the length of the assigned readings, absorb so large a share of study time that the danger of slighting other subjects is great unless recreational activities are curtailed. A student lacking the skills necessary to satisfactory performance in a college course inevitably faces the need of expending a great amount of time in maintaining normal progress. Thus, inadequate training in mathematics prevents acceptable work in physics and chemistry, just as ignorance of the rudiments of grammar handicaps the writing of creditable compositions. By dint of hard work, fundamental skills can be acquired.

---

1. Accomplishment in one subject is seriously hampered according to the writer by
    a. outside work
    b. lack of related skills
    c. ignorance of physics
    d. too much recreation
    e. lack of normal progress
2. The passage is written for
    a. brilliant students
    b. conscientious students
    c. students carrying a heavy outside load
    d. slow students
    e. students weak in elementary skills
3. Lack of basic grammar handicaps, according to this writer
    a. correct speech
    b. work in physics
    c. good writing
    d. effective social relations
    e. accomplishment in chemistry

II

Most successes in prophecy are remembered, failures forgotten. The stock answer is that the stars are never wrong, but that astrologers may be, which is admittedly an argument to end all arguments. The richest field for the study of credulity lies in medical quackery whether it be a movie star who attributes her perfect teeth to some particular brand of toothpaste, or a distinguished foreign physician who recommends yeast for a specific ailment in such a way as to allow the advertiser to seduce the public into believing that a cure for infections must be a medicine for all ills.

1. According to this paragraph a good field for the study of quackery is to be found in
    a. the movies
    b. prophesy
    c. public debate
    d. astrology
    e. advertising
2. The passage is an argument for
    a. the taking of yeast
    b. less advertising
    c. less will to believe
    d. more will to believe
    e. not visiting astrologers.
3. The words "distinguished foreign" refer to
    a. physicians
    b. people in general
    c. medicines
    d. infections
    e. astrologers

III

The story tells how Enmerkar, employing a "modern" diplomatic approach, sent an "ambassador" to the Lord of Aratta and suggested surrender. The latter, the epic continued, said he was the protege of Inanna, the great Sumerian goddess of love and war, and thus diplomatically scorned the proposal. Enmerkar then unleashed an ingenious psychological "war of nerves" by sending threats to the Lord of Aratta and his people. He entwined the threats with favors such as supplies of food for the people. The basic characteristics of man's society have remained relatively unchanged since three thousand years before the days of Christ, Julius Caesar, or the Roman Empire.

---

1. According to this passage a war of nerves is
   a. a means devised by the Romans
   b. to be found in the time of Christ
   c. a method originated by Enmerkar
   d. a very ancient practice
   e. technique developed by Hitler
2. The purpose of the writer is to
   a. show historical continuity
   b. warn
   c. tell a story
   d. amuse us
   e. add knowledge
3. The "Lord of Aratta" was
   a. not named
   b. a Sumerian goddess
   c. Enmerkar
   d. Inanna
   e. an ambassodar
4. We are told that a war of nerves consists of
   a. sending ambassadors
   b. supplies of food
   c. making alliances
   d. psychological warfare
   e. combined threats and favors
5. The writer develops the conclusion drawn in the last sentence through offering
   a. the answers to certain questions
   b. the effect of certain causes
   c. concepts
   d. examples
   e. arguments

IV

Economic thought in the eighteenth century was both stimulated and transformed by the new wealth. The prevailing economic doctrines at the beginning of that era were those of the mercantilist school, which upheld the omnipotence of the state and the subordination of the individual to it. Ever since the advent of the Tudors, in economics as in politics, the movement was away from the small unit, the guild and the town, and toward intensified competition along national lines. The prosperity of a country was thought to depend upon governmental regulation of trade and industry. Wealth was thought of in terms of actual gold and silver.

---

1. According to this passage one would be correct if he argued that
    a. political philosophy is constant
    b. trade and industry have always been free
    c. there was no shift from small units
    d. governmental regulation is not an innovation
    e. economic thought is not subjected to the influence of events
2. A word not used in the paragraph is
    a. regulation
    b. subordination
    c. omnipotence
    d. democracy
    e. intensified

V

In his search for facts and principles, it is the aim of the scientist to proceed in a purposeful, orderly, and logical manner. He must first be adequately prepared to proceed. The modern chemist, physicist, or biologist must serve an arduous apprenticeship in his particular branch of science. So early is this specialization begun that too often no broad basis is laid and no really adequate foundation is afforded for a lifetime of effective citizenship or even of scientific research. Many a biologist, for example, has been balked of high attainment through inadequate training in chemistry or in physics.

1. This writer states that the sciences are
   a. purposeful
   b. logical
   c. interrelated
   d. inadequate
   e. orderly
2. Faults are found by the writer in the demands of
   a. research
   b. arduousness
   c. specialization
   d. logic
   e. chemistry
3. The reader may infer from the passage that sometimes scientific training
   a. builds effective citizens
   b. teaches respect for facts
   c. defeats its own ends
   d. leads to research
   e. teaches orderly procedure

VI

Some readily observable evidences of growth are increase in the size and number of organs. A gardener thinking in terms of some particular crop, measures growth chiefly by the development of the particular part or parts of the plant yielding the products in which he is interested. A forester measures growth primarily in respect to the height and diameter of his trees. A student of botany may be concerned with the grosser evidences and measurements of growth; but for a broad grasp of the growth concept it is essential also to consider the variety of growth phenomena, the manner in which growth proceeds, the detailed mechanism of growth processes, the environmental factors.

1. Which of the following is not mentioned in the passage?
    a. gardeners
    b. organs
    c. students
    d. biologists
    e. foresters
2. The theme of the passage is developed by
    a. scientific analysis
    b. comparison
    c. showing cause and effect
    d. deduction
    e. elimination
3. A gardener is interested chiefly in
    a. measures
    b. phenomena
    c. environmental factors
    d. products
    e. mechanisms

VII

Greatness in prose will always be on a lower level than greatness in poetry. We have seen that what poetry can achieve by the magic of emotive language, imagery, rhythm, and the other devices of communication, prose must attain by the more pedestrian means of cumulative detail, and by exposition, example, illustration, and the other devices of rhetoric and logic. Prose can never hope to speak so intimately to the imagination of the reader, though in its more impassioned passages it may break over into the domain of poetry. Yet, in its secret of greatness, prose is never essentially different from poetry.

---

1. According to this paragraph
   a. prose is essentially different from poetry
   b. prose does not have as much appeal as poetry
   c. prose has as much appeal as poetry
   d. prose has a greater appeal than poetry
   e. no comparison can be made between prose and poetry
2. According to this passage great poetry is essentially different from great prose.
   a. Yes
   b. No
   c. Passage does not say

VIII

To what purpose, April, do you return again?
Beauty is not enough
You can no longer quiet me with the redness
Of little leaves opening stickily.
I know what I know.
The sun is hot on my neck as I observe
The spikes of the crocus.
The smell of the earth is good.
It is apparent that there is no death.
But what does that signify?
Life in itself is nothing
An empty cup, a flight of uncarpeted stairs.
It is not enough that yearly, down this hill,
April
Comes like an idiot, babbling, and strewing flowers.

---

1. The writer is best characterized as being
   a. sorrowful
   b. disillusioned
   c. angry
   d. optimistic
   e. doubting
2. The season attempts to quiet the writer with its
   a. opening leaves
   b. redness
   c. sunshine
   d. babbling
   e. beauty
3. The month is compared to
   a. an idiot
   b. an empty cup
   c. the earth smell
   d. crocus spikes
   e. uncarpeted stairs
4. The line "it is apparent that there is no death" refers most exactly to
   a. the warm sun
   b. the growing flowers
   c. life's emptiness
   d. the red leaves
   e. the recurrent seasons
5. The theme of the poem is that the writer does not intend to allow the month to make her feel that life is enjoyable.
   a. Yes
   b. No
   c. Passage does not say

APPENDIX B

ACKNOWLEDGEMENTS TO PUBLISHERS

The reading passages of the Freshman Oral Reading Test were adapted from the following sources with the permission of the publishers cited therewith.

1. Passage I, Form A and Passage I, Form B, were adapted from Charles Bird's *Effective study Habits* published by D. Appleton Century Company, New York.

2. Passage II, Form A, and Passage II, Form B, were adapted from Wayland Vaughan's *General Psychology* published by the Odyssey Press, New York.

3. Passage III, Form A, and Passage III, Form B, were adapted from the *New York Times* for March 2, 1947. Passage VIII, Form A, was adapted from the *New York Times* for March 30, 1947. These adaptations were made with the permission of the managing editor and with the permission of the editor of the *New York Times Book Review*.

4. Passage IV, Form A, and Passage IV, Form B, were adapted from *A History of England and the British Empire* by W. P. Hall, R. G. Albion, and J. B. Pope. Ginn and Company, New York, are the publishers.

5. Passage V, Form A, and Passage V, Form B, were adapted from E. E. Stanford's *Man and the Living World* published by the Macmillan Company, New York.

6. Passage VI, Form A, and Passage VI, Form B, were adapted from *A Textbook of General Botany* by G. M. Smith, R. E. Evans, B. M. Duggar, G. L. Bryan, and C. E. Allen. The Macmillan Company, New York, are the publishers.

7. Passage VII, Form A, and Passage VII, and VIII, Form B, were adapted from Philo M. Buck's *Literary Criticism, A Study of Values in Literature* published by Harper and Brothers, New York.

REFERENCES

1. Camp, Cordelia. An oral reading program in the elementary grades. *Elem. Sch. J.*, 1932, **32**, 435-41.
2. Courtis, S. A. An analysis of reading ability. *J. educ. Psychol.*, 1921, **4**, 287-93.
3. Davis, F. B. Item analysis data: their computation, interpretation, and use in test construction. *Harvard Educ. Papers, No. 2*, 1946.
4. Eurich, A. C. The relation of speed of reading to comprehension. *Sch. & Soc.*, 1930, **32**, 404-06.
5. Gilmore, J. V. *The relationship between oral reading habits and silent reading comprehension.* Unpublished doctor's thesis. Harvard Univ., 1947.
6. Gray, C. T. The purposes and value of oral reading in the intermediate and upper grades of the elementary school. *Elem. Sch. J.*, 1929, **29**, 335-43.
7. Kelley, T. L. *Fundamentals of statistics.* Cambridge, Mass.: Harvard Univ. Press, 1947.
8. McCullough, C. M., Strang, R. M., & Traxler, A. E. *Problems in the improvement of reading.* New York: McGraw-Hill, 1946.
9. Monroe, M. Methods for diagnosis and treatment of cases of reading disability. *Genetic Psychol. Monogr., No. 4*, 1928.
10. Paul, Vera A. The relation of oral reading to remedial reading in elementary and secondary schools. *Quarterly J. Speech*, 1943, **29**, 217-22.
11. Payne, Cassie S. The classification of errors in oral reading. *Elem. Sch. J.*, 1930, **21**, 142-46.
12. Peters, C. C., & Van Voorhis, W. R. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill, 1940.
13. Sangren, P. V. Methods of diagnosis in reading. *Elem. Eng. Rev.*, 1930, **7**, 105-17.
14. Strang, Ruth. *Examiner's diagnostic reading record.* New York: Teach. Coll. Bureau of Publications, 1939.
15. Swanson, D. E. Common elements in silent and oral reading. Studies in the psychology of reading I. Univ. Iowa studies in psychol., No. 21. *Psychol. Monogr.*, 1937, **48**, 36-59.
16. Thorndike, E. L. Reading as reasoning—a study of mistakes in paragraph reading. *J. educ. Psychol.*, 1917, **8**, 323-32.
17. Tinker, M. A. Rate of work in reading performance as measured by standardized tests. *J. educ. Psychol.*, 1945, **36**, 217-28.
18. Triggs, F. O., & Others. *Diagnostic reading tests.* New York: Educational Records Bureau, 1947.

VOLUME 64
NUMBER 3

WHOLE NO. 309
1950

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

Rorschach Responses Related to Vocational Interests and Job Satisfaction

By

SOLIS L. KATES

Michigan State College
East Lansing, Michigan

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Faculty of Philosophy, Columbia University, 1948.

Accepted for publication, August 31, 1949

Price $1.00

Published by

THE AMERICAN PSYCHOLOGICAL ASSOCIATION
1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

Copyright, 1950, by the
American Psychological Association

ACKNOWLEDGMENTS

To Professor Donald E. Super, I wish to express my deep appreciation for his keen interest, his valuable criticisms and his helpful suggestions throughout the study.

To Professor Laurance F. Shaffer, I am indebted for his invaluable counsel and constant encouragement.

To Professor Robert L. Thorndike, I am grateful for his assistance with the statistical work of the study and for constructive suggestions in presenting the data and conclusions.

Dr. Ruth L. Munroe gave many helpful suggestions in connection with the Inspection Technique. To Miss C. Mildred Ceres for her willing and generous aid, I am very appreciative. I am thankful to Drs. Bernard Locke and Peter J. Napoli who assisted in securing the cooperation of the subjects.

Direct personal acknowledgment, unfortunately, cannot be made here to the anonymous subjects who so willingly cooperated. I also wish to express my thanks to Mrs. Doris Frankel and Mr. Irving A. Tucker for their help with the clerical work.

Finally, my wife made this study possible by her deep devotion and willingness to make many personal sacrifices.

S. L. K.

New York City
1948

CONTENTS

I. The Problem and Aims of the Study 1
  Theories Respecting the Relationship Between Job Satisfaction, Personality, and Vocational Interests 1
  The Problem: Aims of the Study 2
II. Survey of the Literature 4
  Previous Investigations Related to This Study 4
  Previous Investigations Related to the Tests Used in This Study 4
III. Selection of Subjects and Tests and Methods of Gathering Data 6
  Selection of Subjects 6
  The Tests 7
IV. Results 10
  Job Satisfaction Measurement 10
  Strong Vocational Interest Scores 10
  Job Satisfaction and Rorschach Responses 12
  Vocational Interests and Rorschach Responses 15
  Intercorrelations Among Certain Rorschach Factors 17
  Rorschach Responses of Routine Clerks 18
V. Interpretation of Results 20
  Job Satisfaction 20
  Job Satisfaction and Vocational Interests 21
  Rorschach Adjustment 23
  Job Satisfaction and Rorschach Adjustment 23
  Vocational Interests and Rorschach Adjustment 24
  Vocational Interests and Rorschach Responses 25
  Rorschach Responses of Routine Clerks 25
  Munroe Inspection List and Davidson List of Signs 25
  Prediction of Clerical Interests from Munroe List Categories 26
  Vocational Interests and Rorschach Responses: Theoretical Implications 26
  Further Research Suggestions 26
VI. Summary and Conclusions 28
Appendix A. Job Satisfaction Blank 31
Appendix B. Arbitrary Weights, and Entries Considered Correct for Biserial and Tetrachoric Correlations (Munroe Inspection List Categories) 32
Bibliography 33

CHAPTER I

THE PROBLEM AND AIMS OF THE STUDY

THEORIES RESPECTING THE RELATIONSHIP BETWEEN JOB SATISFACTION, PERSONALITY, AND VOCATIONAL INTERESTS

THE job, in our culture, has been variously described as a symbol of status, a source of self-respect, a fundamental basis for the formulation of values, and an important contributor to mental health. A modicum of satisfaction has been recognized as essential to the individual as one of the products of his work. Work satisfaction should be assessed in connection with interests and personality so that the contributions of each may receive proper emphasis. The relation of work satisfaction to the degree of personality adjustment and vocational interest has been speculated upon. In counseling situations, work dissatisfaction has often been ascribed to personality maladjustment, lack of appropriate vocational interests, or both. This point of view was demonstrated in a book devoted to this topic, with its principal conclusion that:

> Much of vocational maladjustment is merely an aspect of the still bigger problem of emotional maladjustment. The reason why this fact has not been adequately recognized by those interested in vocational adjustment is simple; just because he does not know, the individual seldom assigns the true cause as the factor underlying his vocational maladjustment. He is dissatisfied and simply because his work fails to give him the satisfaction he is seeking, he thinks his work is to blame (14, p. 21).

The same authors asserted that nine times out of ten the individual, who is more or less constantly in a state of excessive emotional tension, is vocationally maladjusted (14). Another study hypothesized that deep-seated personality maladjustment might cause individuals to be vocationally dissatisfied, despite their possession of vocational interests similar to those successfully engaged in the same occupation (50). Kimball Young has asserted that the emotionally maladjusted worker's inability

> to adapt himself to the demands of the world reflect back into himself in terms of thwarted wishes, unhappiness, and feelings of failure—a discrepancy between his achievement and his aspiration. And naturally such a person carries these attitudes or "emotional complexes" into new situations. Usually, the individual does not know why he is unhappy or dissatisfied or unsuccessful, and it is not uncommon for him to project blame for his maladjustment upon his job and the working situation (65, pp. 615-616).

Shaffer has averred that only lately has recognition been given to the viewpoint that much vocational dissatisfaction is due to the existence of general personal maladjustment in individuals (51, pp. 516-517). Quite recently, this topic has been discussed as follows: "The individual who is satisfied and adjusted with respect to his working situation is likely to be one who is well adjusted in his family, social and community life" (16, p. 448).

While not denying that personality maladjustment may cause job dissatisfaction, another view has been that either might cause the other and what would remedy one should remedy the other (25). Goodwin Watson's results did not support the contention that poor morale on the job was largely due to personal maladjustment (61). Kornhauser's opinion is that few workers were irremediably of low morale because of their emotional maladjustment, declaring that external variations are still influential (29).

Very little has appeared in the literature relating work satisfaction to measured interests. As Super has pointed out:

> In spite of the widespread acceptance of the hypothesis concerning the significance of vocational interests for personal and vocational adjustment, surprisingly little research has been done in the prediction of satisfaction (54, p. 364).

Strong utilized vocational satisfaction as one of the criteria of validity of his Vocational Interest Blank, stating "Its validity should be measured in terms of satisfaction not while the person is preparing for the work but later on while he is engaged therein" (52, p. 384). Darley reported that the "fulfillment of the interest type (on the Strong Blank) represented a personally satisfying adjustment for the individual", and that this type of satisfaction if realized on the job perpetuated itself and was an important phase of adult adjustment (9, p. 51).

The responses of the individual in terms of liking, disliking or indifference to the many items on an interest inventory are expressions of his drives and goals and give evidence of how he conceives of himself in relation to these (7, 50). Vocational interests have been conceived as distinct but integral parts of the basic personality structure, susceptible to the same influences (4).

THE PROBLEM: AIMS OF THE STUDY

This study investigated the vocational interests, Rorschach responses, and job satisfaction of a group of workers employed by the Federal government and engaged in routine clerical work. It attempted to discover whether those who professed greater satisfaction with their work could be distinguished on the clerical scale of the Strong Vocational Interest Blank from those who indicated less satisfaction. At the same time, the question was raised whether the more satisfied clerks could be distinguished from the less satisfied by the number of Rorschach signs of adjustment as evaluated by two objective scoring techniques, the Davidson List of Adjustment Signs and the Munroe Inspection Technique. Information was also sought on the relationship of the clerical scores on the Strong Vocational Interest Blank to the number of Rorschach signs of adjustment and maladjustment as appraised by the Davidson List and the Munroe Inspection Technique. The study included an attempt to ascertain which, if any, of the specific Rorschach responses were related to work satisfaction. Finally, there was an exploration of the relationship of Rorschach responses of routine clerks to their scores on the Strong Vocational Interest Blank, clerical scale.

Succinctly stated, this study undertook to discover the interrelationships among such factors as job satisfaction, vocational interests, and Rorschach signs of maladjustment. Those Rorschach responses, if any, that were significantly related to job satisfaction and vocational interests, were segregated.

The following hypotheses were set up to be tested:

1. Job satisfaction in clerical work bears a significant positive relationship to scores on the clerical scale of the Strong Vocational Interest Blank. Those clerks who possess interests similar to those of successful office workers tend to be more satisfied.

2. Job satisfaction in clerical work is not related to Rorschach signs of maladjustment. The degree of job satisfaction of clerical workers is independent of the number of Rorschach signs of maladjustment.

3. The work dissatisfaction of routine clerical workers who possess the interests

of successful office workers is not related to the number of Rorschach signs of maladjustment.

4. Certain Rorschach responses, evidenced in the records of routine clerical workers, are significantly associated with their possession of vocational interests similar to those of successful office workers.

ture relating work satisfaction to measured interests. As Super has pointed out:

> In spite of the widespread acceptance of the hypothesis concerning the significance of vocational interests for personal and vocational adjustment, surprisingly little research has been done in the prediction of satisfaction (54, p. 364).

Strong utilized vocational satisfaction as one of the criteria of validity of his Vocational Interest Blank, stating "Its validity should be measured in terms of satisfaction not while the person is preparing for the work but later on while he is engaged therein" (52, p. 384). Darley reported that the "fulfillment of the interest type (on the Strong Blank) represented a personally satisfying adjustment for the individual", and that this type of satisfaction if realized on the job perpetuated itself and was an important phase of adult adjustment (9, p. 51).

The responses of the individual in terms of liking, disliking or indifference to the many items on an interest inventory are expressions of his drives and goals and give evidence of how he conceives of himself in relation to these (7, 50). Vocational interests have been conceived as distinct but integral parts of the basic personality structure, susceptible to the same influences (4).

THE PROBLEM: AIMS OF THE STUDY

This study investigated the vocational interests, Rorschach responses, and job satisfaction of a group of workers employed by the Federal government and engaged in routine clerical work. It attempted to discover whether those who professed greater satisfaction with their work could be distinguished on the clerical scale of the Strong Vocational Interest Blank from those who indicated less satisfaction. At the same time, the question was raised whether the more satisfied clerks could be distinguished from the less satisfied by the number of Rorschach signs of adjustment as evaluated by two objective scoring techniques, the Davidson List of Adjustment Signs and the Munroe Inspection Technique. Information was also sought on the relationship of the clerical scores on the Strong Vocational Interest Blank to the number of Rorschach signs of adjustment and maladjustment as appraised by the Davidson List and the Munroe Inspection Technique. The study included an attempt to ascertain which, if any, of the specific Rorschach responses were related to work satisfaction. Finally, there was an exploration of the relationship of Rorschach responses of routine clerks to their scores on the Strong Vocational Interest Blank, clerical scale.

Succinctly stated, this study undertook to discover the interrelationships among such factors as job satisfaction, vocational interests, and Rorschach signs of maladjustment. Those Rorschach responses, if any, that were significantly related to job satisfaction and vocational interests, were segregated.

The following hypotheses were set up to be tested:

1. Job satisfaction in clerical work bears a significant positive relationship to scores on the clerical scale of the Strong Vocational Interest Blank. Those clerks who possess interests similar to those of successful office workers tend to be more satisfied.

2. Job satisfaction in clerical work is not related to Rorschach signs of maladjustment. The degree of job satisfaction of clerical workers is independent of the number of Rorschach signs of maladjustment.

3. The work dissatisfaction of routine clerical workers who possess the interests

of successful office workers is not related to the number of Rorschach signs of maladjustment.

4. Certain Rorschach responses, evidenced in the records of routine clerical workers, are significantly associated with their possession of vocational interests similar to those of successful office workers.

CHAPTER II

SURVEY OF THE LITERATURE

PREVIOUS INVESTIGATIONS RELATED TO THIS STUDY

THE literature relative to job satisfaction, vocational interests, personality and Rorschach responses, when each of these is considered independently, is so extensive that it would be impractical to present a complete survey. Readers seeking fuller discussions on job satisfaction are referred to Hoppock and collaborators (18, 25); on the Strong Vocational Interest Blank, to Strong (52) and Carter (8); on the Rorschach Test, to Klopfer (28), Hertz (20, 23), and Beck (1).

Job satisfaction has been discovered to be associated with: occupational level (53), a sense of fulfillment in constructive work (13), and the fulfillment of the social demands of the individual (48). The evidence has been conflicting on the relationship of age to job satisfaction (30). No one single factor or group of factors operated alone in creating satisfaction, but it has been recognized that workers reacted to their understanding of the whole situation (29).

Darley apparently confirmed his hypothesis that interest types are related to the development of personality in his study of 1000 subjects (9). The significance of some pertinent investigations of the relationship between vocational interests and personality has been undermined: (*a*) by the use of subjects who were not working, and (*b*) by the inclusion of occupations, under one interest type, that are probably dissimilar in many traits, e.g. artist and physicist, and correlating the scores on the interest type scale with personality measures.

Dodge concluded that the more successful clerical workers in their self-evaluations seem to be less ready to assume responsibility, less at ease in social contacts, and more dependent upon others than the less successful workers (11, 12), within routine work situations. There was a marked tendency for psychoneurotic ex-soldiers to drift into clerical work after failing to adjust in other types of employment (44).

Occupational dissatisfaction has been found to be associated with a lack of primary interest pattern in the current occupation as measured by Strong's Blank (50). Eighty-two percent of 241 college graduates who entered occupations for which high scores were earned on the Strong Vocational Interest Blank were satisfied (34).

PREVIOUS INVESTIGATIONS RELATED TO THE TESTS USED IN THIS STUDY

Rorschach Test

While the Rorschach method has become one of the most popular instruments for assessing the whole personality, much doubt has been cast upon its validity and reliability as a personality test. Experimental evidence of the validity of the Rorschach method has been gained by giving this test under experimentally altered conditions and demonstrating its extreme sensitivity to changing conditions, attitudes, or emotional states (32). Others have compared the Rorschach results with outside criteria and have demonstrated marked agreement (6, 40). The case for the validity of the Rorschach Group Test rests upon the evidence for the individual test (23, 24, 33). The reliability of the Rorschach Test has been studied using both test-retest and split-half methods, with results permitting conflicting interpretations (15, 19, 20, 27, 55, 59).

The psychological significance of the Rorschach movement (M, FM) and the color (FC, CF, C) responses has not been subjected to careful validation, and these responses or categories have unfortunately been accepted and used by many Rorschach workers despite the lack of demonstrated validity (22). There has been essential agreement in the meanings assigned to the Rorschach human action (M) response (22, 28, 49). The results of studies that correlated the Rorschach Test responses with personality inventories are in disagreement with the concept of the human action (M) response as an indicator of maturity (21, 60). The validity of the human action (M) response has been purportedly demonstrated by hypnosis (31), while some reliability coefficients for this Rorschach response have been ascertained to be .62 (59), .75 (20), and .52 (27).

Again, there has been essential agreement on the psychological significance of the Rorschach animal action (FM) response (28), and the Rorschach color (FC, CF, C) responses (6, 28, 49). There has been some evidence that emotional stability and adaptability varied directly with the excess of definite form-color (FC) responses over indefinite (CF) and exclusive color (C) responses (21). There has been no confirmation that more human action (M) than animal action (FM) responses imply maturity (21, 60). Reliability of color responses in the Rorschach

Test has been ascertained as .30 (59), .81 (20), .69 (15), and .68 (55). No apparent reason seemed to explain these discrepancies.

The Munroe Inspection Technique

The Munroe Inspection Technique (40) is an objective method of evaluating Rorschach responses that yields numerical ratings indicative of the degree to which these responses deviate from what has been empirically determined to be adequate. By means of the Munroe Inspection scores, Munroe was able to discriminate with a high degree of accuracy between students who were poorly adjusted and those who were adequately adjusted (40). Roe's investigation demonstrated that while the usual methods of Rorschach Test analysis were unsuccessful in picking out the excessive drinkers, the application of the Munroe Inspection Technique brought out clear differences (45). For 250 college and high school students, there was good correspondence between the Munroe Inspection score and the degree of maladjustment as determined by external indices (38). Fifteen "problem" students selected independently of the Rorschach Test had an average Munroe Inspection score of 7.9 with a range of 5-10. A group of students selected as unusually well adjusted, had an average score of 2.6 on the Munroe Inspection List with a range of 1-4 (36).

The Davidson List of Signs

The Davidson List of Signs was established on a group of 102 superior children (10). Margulies, using the Davidson List of Signs, was able to distinguish between scholastically successful and scholastically unsuccessful boys of equivalent intelligence (35).

The Strong Vocational Interest Blank

The Strong Vocational Interest Blank has been called the "outstanding device of its kind" (8 p. 14). For the appraisal of a vocational interest test, Strong asserted that he could think of no better criterion than that of satisfaction enduring over a period of time (52). Carter furthermore stated that the real purpose of the vocational interest inventory is to predict vocational satisfaction and to indicate vocational preference (8). Strong has stated that those who are satisfactorily adjusted in an occupation may be differentiated by the test from those who are not satisfactorily adjusted and from those not engaged in the occupation. Due to the inadequacy of measures of work satisfaction, Strong has assumed that continuance in an occupation is a measure of satisfactory adjustment, and if a person progressed on his job and was reasonably satisfied he was well adjusted vocationally (52). A test-retest correlation of .59 following an interval of three years was found for the Strong office worker key (58). Strong's reliability coefficient after a five-year interval for the office worker scale was .74 while the odd-even technique yielded a coefficient of .88 (52).

The Job Satisfaction Blank

The blank (Appendix A) used in this investigation was almost identical to that devised by Hoppock for his study of job satisfaction (25). This blank required the subject to evaluate his feelings about the job as a whole. Having the subject do so placed the responsibility upon him of weighing all the job factors. Morale or satisfaction with work was exclusively the sum of attitudes toward work situations and it was not known which of several aspects of a job denoted morale (63). A forty-four item questionnaire, sampling almost every area of the subject's work environment, produced substantially similar results in terms of job satisfaction as did one item requesting the subject to evaluate his feelings of satisfaction with the job (17).

The split-half reliability of the job satisfaction blank was .93 after correction by the Spearman-Brown formula from a raw correlation of .87 (25). Berdie ascertained that the reliability coefficient, also split-half, was .97 (3). The writer carried out a study of 60 subjects who answered the job satisfaction blank at intervals of three weeks. A test-retest reliability coefficient of .90 was obtained.

CHAPTER III

SELECTION OF SUBJECTS AND TESTS AND METHODS OF GATHERING DATA

SELECTION OF SUBJECTS

THE subjects for the study were chosen with a view to keeping to a minimum the variations in factors, other than the experimental ones, which might affect job satisfaction. The choices which were made, and the reasons for them, were as follows:

1. A group of men was used, because earning a living more clearly represents a permanent activity for them.
2. All subjects came from the same occupational group so as to eliminate the effect upon job satisfaction of differences between occupations.
3. All subjects enjoyed, presumably, the same general physical working conditions so that dissimilarities in conditions that might influence job satisfaction were held to a minimum.
4. All subjects earned annual salaries that did not vary too greatly so that the effect of salary differences upon job satisfaction was minimized.
5. The subjects, for practical considerations, were engaged in an occupation for which a scale on the Strong Vocational Interest Blank had been established.

One hundred male routine clerical workers employed in one establishment of the United States Government were selected.

No selective factor was apparently operative in the choice of these subjects. Virtually all the male office workers within one administrative department were asked by the department head to report for the testing required for this study. There was no coercion. All those who wished participated. Many factors operated to bring about 100 percent cooperation. First and most important, neither the subject nor his test results could be identified by the examiner. The subjects drew their test blanks in random order from the stack. Only numbers appeared on the papers. The appeal of one of the tests was great, due to the publicity given by many popular magazines and a recent moving picture. Others reported because it provided a welcome adventure during working hours away from their work duties. And finally, the spectacle of many of their work intimates eagerly volunteering impelled many others to present themselves. Sufficient time off from work was permitted so that there was no need for haste in taking the examination. None of those who had been employed for less than six months were included.

The subjects averaged 29.7 years of age and had completed 12.1 years of schooling, as compared with Strong's criterion group for office workers which averaged 33.2 years of age and had attained a mean of 11.5 years of education (52). Only one routine clerk had completed no more than elementary school while four subjects had finished four years of college. Two clerks had one year of high school, eight had two years, thirteen had three years, while fifty-one completed four years of high school. There were eleven with one year, seven with two years, and three with three years of college work.

Ninety-two percent of the routine clerks were born in this country with 74 percent giving New York City as their birthplace. Only eight other states were represented. Of the eight percent who were of foreign birth, two percent each were born in Russia, Poland, Germany, and the British West Indies. All were citizens of the United States. Thirty-three percent of the subjects' fathers were born in this country. Italy as the fathers' birthplace accounted for 36 percent with Russia having 15 percent. The British West Indies was represented by five percent and Poland had four percent. Norway, Hungary and Germany each showed two percent while Austria had one percent.

Utilizing the occupational scale devised by Beckman (2), the fathers' occupations were classified as follows: Grade I—Unskilled Manual Occupations, Grade II—Semi-Skilled Occupations, Grade III—Skilled Manual and Skilled White Collar Occupations, Grade IV—Sub-Professional, Business and Minor Supervisory Occupations, Grade V—Professional, Managerial, and Executive Occupations. This scale, according to the originator who cited evidence, indicates the rank and the socio-economic prestige attached to a given occupation based on intelligence, skill, education, and training required for its pursuit. The occupations of the fathers of these clerks were distributed among the following grades in this proportion: Five percent were in Grade I, seven percent were in Grade II, 78 percent in Grade III, six percent in Grade IV, and four percent in Grade V. Three of the four in Grade V were musicians. The great majority of the subjects' fathers had been or were engaged in skilled, manual or white collar occupations. The aforementioned facts should provide a framework for understanding the socio-economic background of these subjects and for evaluating results of this study.

The subjects were classified under Federal Civil Service principally as clerks, 10 as CAF-2, and 87 as CAF-3 clerks, with just three in the CAF-4 category. The salary ranges were from $2300 to $2850 per annum with 95 percent of the clerks earning between $2400 and $2700 yearly. Each subject was entitled to salary increments for each eighteen months of satisfactory service. Their duties, as specified by the Civil Service announcement, were clerical. A more detailed description of their duties might afford a more appropriate perspective with which to view the results of this study. Within one sub-group, the clerks were charged with dealing with forms embodying requests for aid, routing these forms, submitting these written requests to the professional staff for attention, and issuing the necessary written authorizations. Another sub-group of clerks maintained personnel records and requisitioned supplies. An additional sub-group devoted themselves to the processing of requests for payment of services furnished by outside agencies. The voucher sub-group maintained the master control cards, governing the reimbursement limitations for use of community resources. The eligibility unit, in which group there were many subjects, determined the eligibility of applicants for assistance from office records, and took the necessary action relating to these applications pertaining to eligibility. Several were engaged in stenography, working for the sub-group administrators. Other subjects dealt with the routine collection of data and statistics. There was great flexibility in the assignment of these subjects to various tasks. As the deputy administrator directly in charge of personnel allocation stated, "If a man does not care for his work or finds it boring, we make adjustments. We transfer him to other types of work. We have never told a man this is your work—you have to do it or else." The kind of work these clerks did was routine, but vital to the functioning of the organization. They had supervisors with whom they conferred if the matter required any judgment or decision not covered by procedure. Typing, filing, sorting, checking through records, stenography, maintaining records, and other related duties were required.

Many of the routine clerks, generally, seemed to like their fellow workers. One clerk, a stenographer, stated that he found the professional people very congenial to deal with. Other clerks agreed with him and indicated that the professional workers treated them as equals. Others stated there was a pretty wholesome relationship with the professional personnel. A few clerks exclaimed about the pettiness of the supervisors and the unfairness of the efficiency ratings. Some echoed one clerk's declaration that "if I want to stick around and get along, I'd better keep quiet." A small number indicated that personal worth was not appreciated. A few felt that they were underpaid for the jobs they performed. One clerk bitterly stated that the other clerks "don't have a mind of their own. They are willing to be led around." Despite a few such adverse indications, many objective signs of favorable work conditions were in evidence, such as an informal atmosphere, low turnover, and few disciplinary actions. Ample opportunity existed for transfer to more congenial work duties and surroundings, if desired. On the whole, it appeared that the clerks performed their routine duties in rather congenial working conditions.

THE TESTS

The Rorschach Test

The Rorschach Test, developed by Herman Rorschach (49), consists of ten standard inkblots, five gray and black and five colored cards. This test has become increasingly popular as an aid to differential diagnosis in clinical situations. It has been used in such fields as anthropology, social case work, sociology, penology, educational and vocational guidance, and personnel selection. It has been assumed by many that the mental procedure which the individual has adopted and the emotional reactions which he has displayed in the Rorschach situation have operated similarly when he met other life situations. In his Rorschach Test reactions he has been thought of as giving a small scale sampling of his characteristic behavior (22). Results have been obtained which, within certain limits, could be subjected to statistical study and analyzed qualitatively to portray the presumed dynamics of personality or behavior (22). One writer has declared that the Rorschach method has resulted in a configurational picture which has revealed the interplay between various major intellectual and emotional factors in the personality of the subject (28). It has been asserted by some that the personality of the individual, revealed in his projections or reactions, can be reconstructed by the experienced Rorschach examiner who can gain knowledge of mental procedures, motives, interests, desires, fantasy life, emotionality, ability, and the like (1, 22, 28); others, on the other hand, have pointed out the need for caution in interpreting Rorschach responses, because of the subjectivity of most of the evidence on which interpretations have been based (22, 66). For this reason, this study has utilized the Rorschach only as an instrument yielding certain objectively measurable indices, which might be related to other variables such as inventoried interest and expressed job satisfaction.

The Munroe Inspection Technique.

As an important outgrowth of her study of the validity of the Rorschach as a "clinical" method, Munroe (40) has developed a special and relatively objective procedure for dealing with Rorschach protocols, which she has termed the *Inspection Technique*. In her preface, Munroe has stated:

> For the single purpose of providing a measure of what—for want of a better word—we have called adjustment it seems probable that the Rorschach can be handled like an ordinary psychological test with the customary attention to set scoring standards and with at least very important reduction in the role played by the judgement of the examiner. Moreover, the measure of adjustment can be expressed quantitatively either as a rating or (probably) as a score obtained by adding up the entries made on a prepared check list (40, p. 6).

The quantitative measure of Rorschach adjustment in this study was obtained by evaluating the Rorschach records of the 100 routine clerks on the basis of the revised Munroe Inspection Technique (37). This check list consisted of the major categories commonly used in Rorschach scoring and for Rorschach interpretation. Each category had fixed limits for what have been clinically considered as "normal" Rorschach responses by Rorschach examiners. Specific numbers of entries were made for each category according to the extent of deviation from these "normal" limits.

Rorschach areas of "adequate" functioning included those responses that were within the empirically established "normal" range of Rorschach responses as shown in each of the categories on the Munroe Inspection List. Clinically, these responses have been found to carry no indications of personality maladjustment. The subject's performance within specific categories or areas was examined to see if it was within the predetermined "normal" range. If within "normal" limits, no entry was made for that specific Munroe category. This meant that the performance of the subject with regard to that Munroe category was "adequate." "Adequate" will have this meaning throughout the study when applied to Rorschach scores.

Rorschach areas of maladjustment embraced those responses that were evaluated as falling outside the empirically established "normal" range of Rorschach responses for each of the separate Munroe Inspection List categories. The degree of deviation of these responses from the predetermined "normal" range was handled by making one, two, or three entries for the particular Munroe Inspection List category. The sum of the entries (or the scores) on the Munroe list for each subject will be referred to throughout the study as the subject's Rorschach signs of maladjustment. Each of the categories in the Munroe Inspection Technique was evaluated in relation to the Rorschach performance of the subject as a whole. A low score on the Munroe Inspection List showed that the subject had given few responses clinically considered to be indicative of maladjustment. A high score on the Munroe Inspection List signified that there were many areas in which he displayed many deviations from the clinically established "normal" range.

The instructions set forth for using the Munroe Inspection Technique were followed in this investigation. Although Munroe emphasized the importance of using clinical judgment in the event that a percentage approached significance in terms of plus, minus, or check scores, the writer decided to adhere strictly to the arithmetical limits for making the number of entries set forth in the directions for the application of the check list. The value of the writer's procedure in terms of making the Munroe Inspection List a more objective measure of Rorschach adjustment was considerable. Also following the Munroe Inspection Technique instructions, another qualified Rorschach examiner and clinical psychologist (Miss C. Mildred Ceres) independently evaluated the protocols. The agreement of the Munroe Inspection scores that she obtained with those obtained by the writer was considerable, the correlation coefficient being .94. It was believed that the merit of this result offset the few advantages that using clinical judgment might have brought.

The Davidson List of Signs.

A high score on this list signified having many Rorschach signs of adjustment while a low score meant the lack of possession of Rorschach signs of adjustment. The Davidson List of Signs (10) was introduced into this investigation to ascertain the extent to which it yielded results similar to those obtained by the Munroe Inspection Technique. It was not as extensive as the Munroe Technique nor did it attempt to distinguish as finely between the differences in quality and quantity of Rorschach responses as did the Munroe. Finally, the Davidson List operated on the principle of the presence or absence of Rorschach signs of adjustment while the Munroe Technique assigned specific values to the degree of presence or absence of specific Rorschach factors of maladjustment.

Procedure Followed for Administering the Rorschach Test.

The procedure followed for administering the Rorschach Test was as follows: Each subject was given a set of Rorschach cards arranged in order, face down, and the instructions suggested by Klopfer (28) were followed, substituting the words "write out" for "tell me." All answers were written in a special booklet with each subject taking as much time as necessary with each of the cards, turning each card if he so desired and was so inclined. After the subject had responded to all the cards, an individual inquiry was conducted with each subject separately so that most of the essential features of the standard individual Rorschach procedure were retained. The subjects were tested in groups of four or five and adequate supervision was provided by the presence of at least two examiners.

The Strong Vocational Interest Blank

The Strong Vocational Interest Blank was utilized to measure the similarity of the vocational interests of the 100 routine office clerks to those held by successful office workers. The standard procedure as prescribed by Strong (52) was followed. This blank consists of 400 items which included names of occupations, school subjects, amusements, activities, peculiarities of people, order of preference of activities, types of famous persons most preferred, choice between two paired activities, and self ratings of present abilities and characteristics. The items were marked according to one's likes, dislikes, indifferences, choices, and preferences. The occupational scale for the general office worker, as for the other occupations, was obtained by contrasting the responses of the criterion group to the 400 items of the blank with those of the men-in-general group, and assigning relative weights to those items that displayed substantial differences. The raw score, obtained by the subject, is the sum of weights he earned on the office worker scale of the Strong Blank, which sum might be positive or negative. The raw score was converted into a standard score which indicated just where the subject placed with respect to the distribution of scores of the criterion group on that scale. The score earned on the general office worker scale was a measure of how similar the subjects' interests were to those of the average successful office workers. A high score signified that the subject possessed interests similar to the successful office worker. A high score was said to be associated with long continued employment and with occupational adjustment and a low score was said to typify the reverse conditions (52). Strong indicated that there was surprising permanence of interests found among adults when vocational interest was defined not as a single selection but as the sum total of many interests that bore in any way upon an occupational career (52). Strong has stated, "men engaged in occupations so far studied have a characteristic pattern of likes and dislikes which differentiates them from men in other occupations" (52, p. 47).

The Job Satisfaction Blank

The job satisfaction blank (Appendix A) was modeled after and differed very little from that devised by Hoppock (25). This blank consisted of four questions, each having seven possible alternatives. These questions required that the subject evaluate himself in connection with how well he liked his job, how he compared with other people in his liking for his job, how he felt about changing his job, and how much of the time he felt satisfied with his job. Job satisfaction had been defined as any combination of circumstances, psychological, physiological, and environmental that made a person state truly, "I am satisfied with my job" (25, p. 47). By having the subject explore and evaluate his feelings in response to the four questions on the aforementioned blank, it was believed that a valid measure of his job satisfaction would be obtained. Rationalization and self-deception may have negated to some extent the subject's own estimate of his job satisfaction. Nevertheless, what the subject thought and felt about his job must necessarily be real despite the origin of such feelings. It was these subjective feelings that this study endeavored to ascertain so that they may be related to the other measures of the individual's functioning—the Strong Vocational Interest Blank, and the Rorschach Test. As Horst has declared, success in an activity may be considered in terms of the satisfaction derived from participation in the activity by the person engaged therein (26). In the majority of prediction studies, however, measures of success have been defined largely in terms of achievement or conformity to norms set by others, rather than in terms of the effect of the activities on the person engaging in them (26). Situational factors, Horst added, must be understood in the way they were defined by the individual himself (26). A true estimate of job satisfaction must come from the subject, notwithstanding the degree of his indulgence in rationalization and falsification.

The score on the job satisfaction blank was the sum of the separate weights for each of the four questions. Each question was given a weight equal to the number of the alternative that was marked by the subject. Hoppock used this method for arriving at job satisfaction scores after discovering that a more complicated and ostensibly more accurate method of assigning weights gave almost identical results, the correlation between the two methods equaling .98 (25). Scores could range from a theoretical low of 4 to a theoretical maximum of 28.

CHAPTER IV

RESULTS

JOB SATISFACTION MEASUREMENT

THE range of scores of routine clerks on the job satisfaction blank was from 7 to 26, signifying that there were subjects who were intensely satisfied with their jobs as well as greatly dissatisfied. The mean of the group was 17.49 points and the standard deviation was 3.80. Table I gives the distribution of the job satisfaction scores. The group median fell at 18, while the mode was 16, the latter score having the top frequency of 17. After arbitrarily fixing 18 as the lowest score indicative of satisfaction, it was noted that 53% of the routine clerks achieved that score or higher. The percentage of the indifferent, those earning scores of 14 to 17, was 33, whereas the dissatisfied, those with scores of 13 and below, constituted 14 percent of the group.

TABLE I

DISTRIBUTION OF JOB SATISFACTION SCORES

Score	Frequency
7 to 10	6
11 to 13	8
14 to 15	10
16 to 17	23
18 to 20	31
21 to 23	18
24 to 26	4

N = 100 Clerks
Mean = 17.5
Standard deviation = 3.8

A certain minimum of intelligence might be prerequisite for satisfaction in any occupation. Uhrbrock has shown that there was a small insignificant negative correlation (—.14) between job satisfaction scores and Otis intelligence scores of 99 clerks (57). Probably, once the minimum intelligence level for an occupation was surpassed, additional increments of intelligence had little effect. The number of years of education of the clerks under current investigation was correlated with their job satisfaction scores and the resulting correlation was —.12, agreeing substantially with Uhrbrock (57). No relationship, hence, for routine clerical workers seemed to exist between job satisfaction and education.

STRONG VOCATIONAL INTEREST SCORES

The scores on the office worker scale of the Strong Blank ranged from 22 to 64. (The standard scores were utilized instead of raw scores due to the occurrence of many negative scores.) The mean score fell at 44.7 with a standard deviation of 9.7 (Table II). This was compared with the mean score of 50 and standard deviation of 10 of the office workers used by Strong as his sample. A difference significant at the 1 percent level existed between the means of the two groups. When the distributions of both groups according to letter ratings were contrasted, a difference significant at the 1 percent level was revealed (Table III). Both means and the letter grade distributions of the two groups were significantly different. Strong's sample of 326 workers, drawn from all parts of the

TABLE II

DISTRIBUTION OF STRONG VOCATIONAL INTEREST CLERICAL SCALE SCORES AND RATINGS

Standard Score	Rating	Frequency
22 to 24	C	2
25 to 29	C+	7
30 to 34	B−	8
35 to 39	B	14
40 to 44	B+	11
45 to 49	A	19
50 to 54	A	27
55 to 59	A	7
60 to 64	A	5

N = 100 Clerks
Mean Score = 44.7
Standard deviation = 9.7

TABLE III

CLERICAL INTEREST RATINGS OF 100 ROUTINE CLERKS AND THE CRITERION GROUP ON THE STRONG BLANK

Group	Letter Ratings			
	A	B+	B	B−, C+, C
Routine Clerks	58	11.0	14.0	17.0
Criterion Group	70	12.3	9.8	7.8

Chi-square = 14.84 3 degrees of freedom
Significant at the 1% level

United States, was composed of 214 office clerks, bookkeepers, stenographers, 92 office managers, and 20 credit managers. Age and educational level of the two groups were about the same or equal. More than one third of Strong's group were subjects charged with the responsibility of making decisions, using initiative, or carrying on administrative duties.

RELATIONSHIP BETWEEN JOB SATISFACTION SCORES AND VOCATIONAL INTEREST SCORES

The correlation coefficient between job satisfaction scores and the scores on the clerical scale of the Strong Vocational Interest Blank was .21. This correlation was not significant at the one percent level but was significant at the five percent level. For this particular group of clerical workers, there was only a slight relationship between degree of satisfaction with a routine clerical job and the score on the clerical scale of the Strong Blank. Vocational interests similar to those of office workers could be a factor contributing to satisfaction with routine clerical work but to a slight degree only. The partial correlation between job satisfaction scores and Strong clerical interest scores with the Munroe Inspection Technique total adjustment scores held constant, was .22. This latter coefficient was not significant even at the five percent level. Hence, there was no evidence to support a relationship in this group of routine clerks between job satisfaction scores and Strong clerical interest scores when Munroe Inspection scores were held constant.

Strong stipulated that the criterion of validity for his test was not a high degree of job satisfaction but "reasonable" satisfaction in contradistinction to indifference and dissatisfaction (52). A threefold division of job satisfaction was made by assigning all scores of 18 and above to the satisfied group, scores of 14 to 17 inclusive to mark indifference, while scores of 13 or below typified dissatisfaction. Following Strong's definition that all those with A and B+ scores were to be considered as possessing interests similar to those of the average successful office worker, the group was dichotomized into those achieving A and B+ scores and those who did not. Table IV shows the distribution and the chi-square value of 2.89, with 2 degrees of freedom. There was no evidence that vocational interest scores were related to job satisfaction ratings. While there was a very slight tendency for numerical values indicative of vocational interest and job satisfaction to be correlated, when these measures were divided into meaningful categories no evidence for them to be related was found. Analysis of variance

TABLE IV

DISTRIBUTION OF THE STRONG CLERICAL INTEREST RATINGS MADE BY 100 CLERKS ACCORDING TO THEIR JOB SATISFACTION RATINGS

Clerical Interest Ratings	Job Satisfaction Ratings		
	Dissatisfied (0 to 13)	Indifferent (14 to 17)	Satisfied (18 to 28)
A and B+ (40 and over)	7	23	39
B, B−, C+, and C (39 and less)	7	10	14

Chi-square = 2.89 2 degrees of freedom
Not significant at the 5% level

also disclosed that there was no evidence for a significant difference between the clerical interest scores of satisfied, indifferent and dissatisfied routine clerks (Table V).

JOB SATISFACTION AND RORSCHACH RESPONSES

Job Satisfaction Scores and Davidson List of Adjustment Signs

Each of the Rorschach protocols of the 100 subjects was scored according to the system utilized by Klopfer (53). The protocols were evaluated on the basis of 16 of the 17 signs on the Davidson List (10). The sign, total number of Rorschach responses equal 20, was eliminated because there was some evidence that the conditions of group Rorschach administration, followed in this study, might have militated against the production of 20 or more responses (24). The mean number of Davidson adjustment signs for the clerical group was 7.91, the standard deviation 2.31. This fell between the mean of 6.85, standard deviation 2.27, made by scholastically unsuccessful boys and the mean of 9.05, standard deviation of 2.70 of successful boys (35). While the routine clerks were not too different from Margulies' (35) total group of children, when compared with Davidson's original group of 102 superior children (10), the routine clerks were shown to have fewer Rorschach signs of adjustment, 7.91 versus 10.31. Tests for the significance of the differences between the means were not made because of the difference in age of the groups and the omission of one sign of the Davidson List.

The coefficient of correlation between the job satisfaction scores and the total number of Davidson signs of Rorschach adjustment was −.16. As this coefficient was not significant, there was no evidence

TABLE V

ANALYSIS OF VARIANCE OF STRONG CLERICAL INTEREST SCORES OF 100 CLERKS DIVIDED ACCORDING TO SATISFIED, INDIFFERENT, AND DISSATISFIED JOB RATINGS

Source	d.f.	Sum of Squares	Mean Squares	F
Between Means	2	497	248.5	
Within Groups	97	8907	91.5	2.7

F is not significant at 5% level

of a relationship between self-evaluation of job satisfaction and the number of Davidson signs of adjustment. Satisfaction with routine clerical work and the Rorschach responses believed by Davidson to be indicative of adjustment did not appear to be related.

Job Satisfaction Scores and Munroe Inspection Scores.

The Munroe Inspection Technique "total adjustment" score indicated the number and extent of the deviations of the Rorschach Test responses of a subject from those of the empirically established Rorschach norms. The higher the score, the more numerous the Rorschach responses considered indicative of maladjustment. The lower the score, the more normal the performance of the subject on the Rorschach. The mean "total adjustment" score of the clerks on the Munroe Inspection Technique was found to be 12.96 with a standard deviation of 4.40. Scores ranged from a low of 2 to a maximum of 23. The mode fell at two scores, 12 and 13 (Table VI). The median score was 13. Eighty college students studied by Munroe (39) earned an average score on the Munroe List of 9.0, with the median about the same, and their range was 2 to 17. Roe found that the average Munroe Inspection score for a group of artists was 10.3, and that the range was 3 to 18 (46). Scientists averaged 7.7 signs, with a range of 1 to 15, and technicians scored a mean of 9.4 signs with a range of 2 to 19 (47). It can be readily seen that the routine clerks now under study have consistently more Rorschach signs of maladjustment than the other groups. These studies did not report sufficient data for tests of significance.

TABLE VI

DISTRIBUTION OF MUNROE INSPECTION TECHNIQUE SCORES FOR 100 CLERKS

Score	Frequency
2 to 4	2
5 to 7	8
8 to 10	18
11 to 12	17
13 to 15	28
16 to 18	16
19 to 21	7
22 to 23	4

Mean = 12.96
Standard deviation = 4.4

It seemed desirable to learn if the Munroe Inspection Technique utilized by another psychologist would yield similar results when applied to the Rorschach protocols. Following the scoring of the Rorschach protocols according to the customary Klopfer method, thirty of them were submitted to a clinical psychologist, experienced in Rorschach examination (Miss C. Mildred Ceres). She independently rechecked the writer's scoring, making alterations and changes, when necessary, and, without knowledge of the writer's Munroe Inspection Technique scores, evaluated each of the protocols based on this technique. The above procedure was followed to appraise the reliability and similarity of the Munroe Inspection Technique scoring rather than the reliability and similarity of the general Rorschach scoring. The coefficient of correlation between the independently computed Munroe Inspection Technique scores was .94, with perfect agreement in 30 percent of the records. Thirty percent of the records differed by 1 entry, 33 percent of the records differed by 2 entries, and seven percent differed by 3. No score differed by more than three. The writer's own reliability coefficient was .96 for the first 20 records rescored after an interval of three months.

The coefficient of correlation between the job satisfaction scores and the Munroe Inspection scores was calculated to be .02, which was not significant. The partial correlation coefficient between these two variables with Strong clerical interest scores held constant (—.06) was not significant. The Rorschach signs of maladjustment, as appraised by the Munroe Inspection Technique, were not related to job satisfaction scores. This

TABLE VII

ANALYSIS OF VARIANCE OF JOB SATISFACTION SCORES OF 100 CLERKS DIVIDED ACCORDING TO RELATIVELY FEW (0 TO 10), MODERATE (11 TO 15), AND MANY (16 AND OVER) RORSCHACH SIGNS OF MALADJUSTMENT ON THE MUNROE INSPECTION LIST

Source	df	Sum of Squares	Mean Square	*F*
Among Means	2	17.47	8.74	
Within Groups	97	1427.53	14.72	0.59

F is not significant at 5% level

result confirmed the result reported above with the Davidson List of Adjustment signs.

The Munroe Inspection Technique scores were divided into three categories, the scores falling between 0 and 10 indicating relatively few Rorschach signs of maladjustment, the scores between 11 and 15 revealing relatively moderate Rorschach signs of maladjustment, and scores of 16 and over representing relatively many Rorschach signs of maladjustment. The corresponding job satisfaction scores were tabulated and the analysis of variance technique was applied to ascertain if the scores were significantly different among the three categories of Rorschach adjustment. No significant differences were discovered to exist among the job satisfaction scores of these three categories of Rorschach adjustment (Table VII). Tests of significance utilizing Chi-square also revealed no relationship between Munroe Inspection scores (Table VIII). Analysis of variance and chi-square procedures were utilized in addition to the Pearson product-moment correlation inasmuch as there might have been some doubt about the Munroe Inspection List as a continuous measure of Rorschach adjustment.

Job Satisfaction Scores and Munroe Inspection List Categories

The following description of results pertains to the separate categories on the Munroe Inspection List and their correlations with job satisfaction scores. In some cases, arbitrary weights have been assigned to specific entries for each category used, in order to calculate product-moment correlations. For the biserial correlations, certain groupings of entries under each category were deemed arbitrarily right and others wrong, but only for purposes of calculation. Appendix B should be consulted for these arbitrary weightings and determinations.

TABLE VIII

JOB SATISFACTION SCORES OF 100 CLERKS SHOWING RORSCHACH SIGNS OF MALADJUSTMENT ON THE MUNROE INSPECTION LIST

Job Satisfaction Scores	Rorschach Signs of Maladjustment		
	Relatively Few (0 to 10)	Relatively Moderate (11 to 15)	Relatively Many (16 and over)
18 to 28	12	17	14
14 to 17	10	13	10
0 to 13	6	5	3

Chi-square = 3.26 4 degrees of freedom
Not significant at 5% level

The coefficients of correlation between job satisfaction scores and the Munroe Inspection List categories have been indicated in Table IX. No one of these correlations was statistically significant. There was no evidence, then, for any relationship between the individual Munroe Inspection List categories and job satisfaction scores of routine clerks.

TABLE IX

CORRELATION COEFFICIENTS BETWEEN JOB SATISFACTION SCORES AND MUNROE INSPECTION LIST CATEGORY ENTRIES

Munroe Inspection List Categories	Pearson Product Moment Correlations	Biserial Correlations
Color to Movement	.02	−.07
CF:FC	−.07	.00
FC	.06	.03
FM:M	−.06	−.11
M	.04	.04
Total Color	—	.10
Total Movement	—	−.13
FK, Fc	—	.05
F%	—	.01
Range	—	.05
Popular	—	.05
Dd	—	−.01
W	—	.13

Job Satisfaction Scores and the Combination of Munroe Inspection Scores with Strong Vocational Interest Scores.

It was desirable to ascertain whether certain combinations of Rorschach responses and vocational interests distinguish routine clerks who had satisfied scores from those whose scores indicated indifference to or dissatisfaction with their work. There was no significant difference in the number of routine clerks with Strong clerical interest scores of 40 or above and Munroe Inspection scores of 16 or above who had job satisfaction scores of 18 or above, on the one hand, and those with the same combination of Strong and Munroe scores who had job satisfaction scores of 17 or below, on the other. There was no evidence that job dissatisfaction or indifference of routine clerks who possessed interests similar to those of successful office workers was associated with Rorschach responses evaluated by the Munroe List as indicative of relatively many signs of maladjustment (Table X).

VOCATIONAL INTERESTS AND RORSCHACH RESPONSES

Vocational Interest Scores and the Davidson List of Adjustment Signs.

The correlation coefficient between the Strong office worker interest scores and the total number of Davidson signs of Rorschach adjustment was −.54, a significant correlation. The possession by routine clerks of vocational interests similar to those of successful office workers was associated with a low number of Rorschach signs of adjustment, as evaluated by the Davidson List.

Vocational Interest Scores and the Munroe Inspection List

The correlation coefficient between the Strong clerical interest scores and the Munroe Inspection scores was .33, which

TABLE X

JOB RATINGS OF 100 CLERKS WHO POSSESSED OR DID NOT POSSESS THE COMBINATION OF STRONG CLERICAL INTEREST SCORES OF 40 OR ABOVE AND MUNROE INSPECTION SCORES OF 16 OR ABOVE

	Satisfied (18 and above)	Dissatisfied and Indifferent (17 and below)
Pattern	12	10
No Pattern	41	37

Chi-square = .027 1 degree of freedom
Not significant at 5% level

TABLE XI

DISTRIBUTION OF STRONG CLERICAL INTEREST RATINGS MADE BY 100 CLERKS SHOWING RORSCHACH SIGNS OF MALADJUSTMENT ON THE MUNROE INSPECTION LIST

Clerical Interest Ratings	Relatively Few (0 to 10)	Relatively Moderate (11–15)	Relatively Many (16 and over)
A and B+ (40 and over)	13	34	22
B, B−, C+, C (39 and less)	15	11	5

Chi-square = 9.52 2 degrees of freedom
Significant at 1% level

was significant at the 1 percent level. The partial correlation coefficient between the same two factors when job satisfaction scores was held constant was the same, .33, again significant. The possession of interests similar to those of successful clerks tended to be associated with Rorschach responses evaluated by the Munroe List as indicative of many signs of maladjustment. This conclusion agreed with that reached in connection with the relationship between the Davidson List of signs of adjustment and the Strong clerical interest scores. The presence of the tendency for high clerical interest scores to be associated with high Munroe Inspection scores was also shown by the chi-square technique (Table XI).

Strong Vocational Interest Scores and Munroe Inspection List Categories.

The following description of results refers to the separate categories on the Munroe Inspection List and their correlations with the Strong clerical interest scores. For the arbitrary weights and the arbitrary assignment of right and wrong answers for purposes of computation, Appendix B should be studied.

The coefficients of correlation between Strong clerical interest scores of routine clerks and the individual Munroe Inspection List categories have been indicated in Table XII.

In the Klopfer scoring system, a response was scored as "FC" if definite form was fused with color, and a "CF" score was assigned where the form was indefinite and the color important (28). The product moment correlation between the entries for the Munroe Inspection List category CF:FC and scores on the clerical interest scale of the Strong Blank was −.41 while the biserial correlation coefficient was −.52. The differences in correlation coefficients here and below, while not great, might be attri-

TABLE XII

CORRELATION COEFFICIENTS BETWEEN STRONG VOCATIONAL INTEREST SCORES AND MUNROE INSPECTION LIST CATEGORY ENTRIES

Munroe Inspection List Categories	Pearson Product Moment	Biserial
Color to Movement	.09	−.06
CF:FC	−.41	−.52
FC	.27	.40
FM:M	−.40	−.50
M	.55	.53
Total Color	—	−.02
Total Movement	—	−.09
FK, Fc	—	.11
F%	—	−.03
Range	—	.16
Popular Responses	—	.23
Dd	—	−.03
W	—	.14

buted to the probable lack of normality resulting from the arbitrary assignment of weights for the calculation of the product moment correlations. The higher the score on the Strong clerical interest scale, the greater was the tendency of routine clerical workers to give preponderantly indefinite form-color (CF) in preference to definite form-color (FC) responses to the Rorschach ink blots.

The product moment correlation between the entries for the Munroe Inspection List category FC and scores on the clerical interest scale of the Strong Blank was .27 which was significant at the 1% level. The biserial correlation was .40. The higher the Strong clerical interest scores of routine workers the greater was the tendency for the less frequent use of definite form-color (FC) responses in the Rorschach records.

A score of FM referred to the spontaneous response by a subject in which he reported perceiving an animal in action or in a live posture, while M applied to responses embodying the concept of a living being in human or human-like action. The product moment correlation coefficient between the entries for the Munroe Inspection List category FM:M and scores on the clerical interest scale of the Strong Blank was −.40 which was significant, while the biserial coefficient was −.50. These correlations revealed that, for routine clerks, high Strong clerical interest scores tended to be associated with a preponderant use of animal action (FM) responses over human action (M) responses.

The product moment correlation coefficient between the entries for the Munroe Inspection List category M and scores on the clerical interest scale of the Strong Blank was .55 while the biserial coefficient was .53. In this sample of routine clerks, high scores on the clerical interest scale of the Strong Blank tended to be associated with infrequent use of human action (M) responses.

There are ten popular responses in the Klopfer system, described in his manual (28). The biserial correlation coefficient between the entries for the Munroe Inspection List category popular responses and scores on the clerical interest scale of the Strong Blank was .23, significant at 5% level. There was a slight but significant relationship between the lack of use of four or more popular responses and the possession by routine clerks of interests similar to those of successful office workers.

INTERCORRELATIONS AMONG CERTAIN RORSCHACH FACTORS

Davidson List of Adjustment Signs and the Munroe Inspection Technique.

The Davidson List of signs of adjustment was introduced to ascertain the extent to which another objective method of assessing Rorschach adjustment would yield results similar to those of the Munroe Inspection Technique. The product moment correlation coefficient for the two measures was −.764, demonstrating a high relationship. A high Munroe score (many Rorschach responses considered signs of maladjustment) strongly tended to be associated with a low Davidson score, representing the absence of Rorschach signs of maladjustment. The correlation coefficients of each measure with job satisfaction and vocational interest scores tended to agree.

The Munroe Inspection List Categories.

For several of the categories of the Munroe Inspection List which correlated significantly with the Strong clerical interest scores, intercorrelations were calculated, using the product-moment and tetrachoric methods. The arbitrary weights and assignments of right and wrong answers are shown in Appendix B. These intercorrelations were essential for arriving at a multiple correlation coefficient for predicting Strong clerical interest scores from a combination of five Rorschach factors (Table XIII). The significant relationships shown in Table XIII indicated that:

1. The greater the preponderance of indefinite form-color (CF) responses to definite form-color (FC) responses, the smaller was the number of definite form-color (FC) responses of routine clerks.

2. The greater the predominance of indefinite form-color (CF) responses over definite form-color (FC) responses, the more infrequent was the use of human action (M) responses by routine clerks.

3. The greater the predominance of animal action (FM) responses to human action (M) responses, the smaller was the number of human action (M) responses.

4. The smaller the ratio of animal action (FM) to human action (M) responses, the greater was the tendency for routine clerks to give four or more popular responses in their Rorschach record.

5. The smaller the number of human action (M) responses, the smaller was the tendency for routine clerks to give four or more popular responses in their Rorschach records.

MULTIPLE CORRELATION BETWEEN STRONG VOCATIONAL INTEREST SCORES AND MUNROE INSPECTION LIST CATEGORIES

Utilizing the following five categories on the Munroe Inspection List, CF:FC, FM:M, FC, M, and Populars, a multiple correlation coefficient of .72 with Strong Clerical Interest Scores was calculated from their tetrachoric correlations. The iterative technique, described in an Army Air Forces research publication edited by R. L. Thorndike (56), was utilized for predicting Strong Interest Scores. This correlation, corrected for shrinkage, was .70.

RORSCHACH RESPONSES OF ROUTINE CLERKS

The average number of the total Rorschach responses of the routine clerks was 21.4, with a standard deviation of 8.7. Lindner's group of 100 prison inmates averaged 20.4 responses (33). Hertzman's 100 college students had 25.1 responses as their mean (24). The mean number of responses for scientists and technicians was 21.0, and 18.2 respectively (47). An unselected group of 34 adults, age range 26 years to 58 years, averaged 30.9 responses (19).

The percentages of the responses distributed among the Rorschach location categories (W, D, d, Dd, and S) for the 100 routine clerks, and the 34 unselected adults tested by Harrower-Erickson and Steiner (19) are shown in Table XIV. The definitions and significance of these Rorschach responses are in Klopfer's manual (28). No significant differences were found between the percentages of Rorschach location responses used by each group.

The percentages of the responses distributed among the Rorschach determinant responses (M, FM, m, k, K, FK, F, Fc, c, C′, FC, CF, C) for the 100 routine clerks and the 34 unselected

TABLE XIII

CORRELATION COEFFICIENTS BETWEEN MUNROE INSPECTION LIST CATEGORIES*

	CF:FC	FM:M	FC	M	Popular Responses
CF:FC		.07 *.02*	−.31 *−.54*	−.32 *−.36*	— *−.07*
FM:M			.00 *.01*	−.41 *−.80*	— *−.22*
FC				.06 *−.03*	— *.14*
M					— *.23*

---

* The italicized coefficients were computed by the tetrachoric method while the others were calculated by the product moment method.

TABLE XIV

Rorschach Location Responses of Two Groups of Adults

Location Responses	Kates: 100 Routine Clerks	Harrower-Erickson and Steiner: 34 Unselected Adults	$\frac{D}{\sigma_D}$
	Percentage	Percentage	
W	42.1	31.8	1.1
D	44.5	48.1	0.4
d	4.7	12.8	1.3
Dd	7.6	5.1	0.5
S	0.2	2.4	0.8

adults are also shown (Table XV). The definitions and significance of these Rorschach determinant responses can be found in Klopfer's manual (28). The only significant difference was in the percentage of animal action (FM) responses used by each group. The other percentages did not reveal significant differences. The routine clerks tended to utilize significantly more animal action (FM) responses than the other adults. This difference might be due to the dissimilarity in the inquiry procedures. Individual inquiries were conducted with the clerical group while the unselected adults received a group inquiry.

TABLE XV

Rorschach Determinant Responses of Two Groups of Adults

Determinant Responses	Kates: 100 Routine Clerks	Harrower-Erickson and Steiner: 34 Unselected Adults	$\frac{D}{\sigma_D}$
	Percentage	Percentage	
M	12.2	9.1	0.5
FM	25.3	5.5	3.5
m	1.7	1.8	0.1
k	1.6	1.1	0.2
K	1.7	1.4	0.1
FK	0.8	2.7	0.7
F	33.7	54.3	1.9
Fc	4.9	4.2	0.2
c	1.3	4.1	0.8
C'	2.5	2.1	0.1
FC	4.4	7.0	0.6
CF	9.3	6.5	0.8
C	0.0	0.2	0.9

CHAPTER V

INTERPRETATION OF RESULTS

JOB SATISFACTION

THE degree of dissatisfaction expressed by these routine clerical workers is noteworthy when contrasted with typical results for adults in general (25). Many of the clerical workers were barely satisfied with their work while a goodly proportion were either dissatisfied or indifferent. Comparison with other investigations (42, 64) indicates that this amount of job dissatisfaction is fairly typical of clerks. Wrenn studied the job satisfaction of college graduates who were engaged in many occupations (64). He discovered that 19 percent were dissatisfied with their jobs in contrast to the 52 percent of the clerical workers in his group who were dissatisfied (64). In another survey, two clerical groups numbering 117 and 102, were satisfied in the proportion of 50 percent and 51 percent respectively (42). The results obtained in the writer's study were strikingly similar to the job satisfaction results described above.

Using the same scale and measuring satisfaction with nursing Nahm ascertained that the mean satisfaction score of 428 subjects was 21.8 and that the standard deviation was 2.86 (41). A heterogeneous sampling of 309 subjects, representative of a small community, averaged 20.01 points (25). (This mean satisfaction score was calculated from two frequency distributions.) The mean satisfaction score of 154 engineering students was 21.06, standard deviation 2.98 (3). The mean satisfaction score of the 100 routine clerks under present investigation appears to be significantly below that reported in the above investigations. In addition, the percentage of satisfied clerks found in this investigation is lower than that found in many other investigations (17, 18, 25).

Workers on higher occupational levels were found to be more satisfied than those on the lower levels (25, 42, 53). It may be presumed that one of the causes for this comparative lack of work satisfaction was the relatively low occupational status of government clerical workers. Associated with this feature were their lower earnings. Probably the higher mean score of the 17 clerks in Hoppock's study (25), that of 21.52, was due to the difference in economic conditions. This last study was conducted during the depression years, 1932-1933. It may be that during depression years workers are happy to hold any job, while in prosperous years they are more dissatisfied when they eye the more lucrative positions. Perhaps the vocational aspiration levels of these clerks were pegged so high as a consequence of the urge to rise in our society that dissatisfaction with what might actually be suitable work ensued unjustifiably.

The type of work demanded of these clerks affords another clue to their dissatisfaction. The work, which was routine, did not demand much individual judgment or initiative, afforded a dearth of social contacts, and required, above all, the ability to take direction. Dodge had shown that the successful routine clerical workers displayed these characteristics. Assuming that the duties of the work did not present any difficulties to the clerks in the matter of personality requirements, then it is surprising that these clerks did show so much dissatisfaction. Either these clerks' personality characteristics were at variance with those demanded by the job, or having personality traits characteristic of routine clerks is not of cardinal importance for the existence of a high degree of job satisfaction.

There is no evidence to support a relationship between the number of years of schooling and the degree of job satisfaction. The number of years of schooling has been used by Wesman to validate his Personnel Classification Test (62). It has been shown by Pond and Bills that years of schooling correlated .50 with intelligence test scores of clerks and other workers (43). Due to the broadening of educational opportunity, there is probably greater correspondence now between years of schooling and intelligence than when Pond and Bills completed their study. Young reported that workers of superior intelligence might be bored by repetitious tasks but on the other hand, they were often able to arrange their work so that it afforded them relief from monotony (65). Assuming that the number of years of schooling was a rough measure of intelligence, the inference might then be made that intelligence and job satisfaction of clerical workers are not related. This inference might be applied to all clerical workers utilizing as evidence Uhrbrock's (57) results and the above mentioned studies

JOB SATISFACTION AND VOCATIONAL INTERESTS

There seemed to be slight evidence that, in routine clerical workers at least, job satisfaction scores were positively related to clerical interest scores on the Strong Blank. The product moment correlation (+.21) was significant at the five percent level. Breaking the scores into meaningful categories as suggested by Strong failed to reveal any evidence of a relationship between letter ratings or numerical scores on the Strong clerical interest scale and being satisfied with, indifferent to, and dissatisfied with routine clerical work. The lack of demonstrated significant relationship between job satisfaction and vocational interests may be a function of the lowered sensitivity of one or both of the tests. Apparently, the clerical scale of the Strong Vocational Interest Blank was of slight value in distinguishing those who were content with their work.

Two additional observations can be made with regard to job satisfaction and vocational interests: the measure utilized in evaluating job satisfaction may have been faulty, and the differences between Strong's criterion group and the 100 routine clerks may have contributed to a lack of a demonstrated relationship. Hoppock (25) and Strong (52) both stated that more satisfactory measures of job satisfaction need to be constructed. In view of the fact that most workers reacted to their situations as a whole, it is felt that the blank utilized was more adequate than many others. It was of importance for this study to obtain a self-evaluation of satisfaction which this blank again provided. The self-evaluation of work contentment could be influenced by many factors extraneous to the work situation, as Kornhauser indicated (29). Some would not be adept at estimating how they actually felt about a job or would not translate their feelings adequately when the job satisfaction questionnaire was laid before them. Some degree of self-deception or rationalization could be operative. While the workers might like clerical work in general, the conditions peculiar to their current employment might so color their thinking that they would evaluate themselves as dissatisfied. Hence, it is conceivable that individuals might receive A or B+ ratings on the clerical scale of the Strong Blank, might like routine clerical work greatly, and yet might express indifference to or dissatisfaction with their job.

It must be recalled that the group upon which Strong standardized his office worker scale was more heterogeneous than this group and earned significantly different scores and ratings on the clerical scale. Approximately one-third of Strong's group were clerks who could be termed administrators having little in common with routine clerical workers. In addition, the groups were not strictly comparable due to the differences in geographic origin, with Strong's group coming from many different parts of the country. In the group of 100 routine clerks, there might have been many who were engaged in clerical work because of economic circumstances despite lack of interests similar to those of clerical workers. Moreover, the possibility exists that the Strong office worker scale was not particularly suitable for routine clerks inasmuch as Strong attempted to establish norms for all types of office workers, including routine clerks. On the other hand, these significant differences should not be construed as impugning the validity of the clerical scale of the Strong Blank. These differences support the conclusion that office workers who originate from one geographic area and who carry on routine work differ significantly in their vocational interests from Strong's criterion group, represented by workers throughout the United States and distributed among all types of clerical workers. Strong has objected to the use of different norms for separate groups within one occupation (52). He pointed out that his norms have been based upon a more representative occupational sampling, usually far more numerous. In treating the subjects' scores on the clerical scale, the point of reference (that of similarity of interests to those of successful office workers, including administrative clerical workers from all parts of the country) must be kept in mind. The thought cannot be lightly dismissed that an occupational scale based on routine clerks as the criterion group might so reshuffle the scores on the Strong Blank that greater correspondence would be exhibited between these scores and job satisfaction. This type of scale would require a population representative of routine clerks throughout the country and would form the basis of another study especially if carried out coincidentally with personality and aptitude tests.

Certain other facts may help in understanding why vocational interest scores were of slight value in distinguishing the more satisfied clerks from the less satisfied ones. Workers who possessed mechanical skills have been found to be more interested in the techniques, procedures, and products of their work than were unskilled workers, who gave predominant importance to the working environment (13). Roethlisberger and Dickson have pointed out, too, that unskilled workers were more attentive to work conditions than supervisors, the latter being more interested in what their work duties demanded than in their environmental conditions (48). Where personal talents and skills are challenged or demanded by a job, it appears likely that correspondence of interests to those of other successful workers in the occupation would be significantly associated with job satisfaction. The satisfaction of achieving something difficult, of expressing one's skills and abilities, and of being able to deal with tasks successfully and constructively has been related to the development of vocational interests. Carter has reasoned that the individual attempted to identify himself with specific occupational groups whose activities he could successfully handle (7). Bordin, essentially agreeing with Carter, has stated that the individual conceived of himself in terms of occupational stereotypes whose requirements he satisfyingly and successfully fulfilled and that a change in this self-concept resulted in a change of vocational interests (5). It is possible, then, that the professional and mechanical occupations would reveal significant relationships between the possession of the related vocational interests and job satisfaction. Further research is desirable, in this matter, to ascertain if such a relationship exists between vocational interests and job satisfaction and to throw light upon this hypothesis. No study, with the possible exception of Lorimer's (34) has yet validated this plausible and logical relationship. But for an occupation such as routine clerical work, the nature of the work was not challenging and little opportunity existed, generally speaking, to achieve something difficult. Not too much correspondence between interests similar to those of office workers and job satisfaction should exist. The intrinsic nature of clerical work contributed relatively little to job satisfaction, when viewed according to the aforementioned standpoint.

It is vital, now, to examine what practical criterion of validity Strong used for his test. He stated:

> Because we had no adequate measure of satisfaction in their work of the men in the follow-up studies to be reported below, we have been forced to content ourselves with assuming that "continuance in an occupation" is a measure of "satisfactory adjustment" (52, p. 387).

Using this criterion, continuance in an occupation, Strong succeeded in validating his blank. Strong declared, however, that continuance in an occupation could not be accepted as the full equivalent of job satisfaction because many dissatisfied workers were not free to leave their employment. It may perhaps be found that the clerical scale of the Strong Blank is more successful in predicting those who will continue in clerical work rather than those who will be satisfied with the occupation. This last statement has not been demonstrated in this study, but these results combined with Strong's, make it a reasonable conclusion. It is essentially a more justifiable conclusion than stating that job satisfaction may be predicted from vocational interest scores, which this study has found to be of questionable accuracy. It can probably be stated that all Strong interest scores could predict continuance in an occupation, with Strong's own data as proof. It is unwarranted to say that men will be satisfied with their jobs if their interests are similar to those of men successfully engaged in such work. This hypothesis still remains unproved even for the professions, despite its plausibility. *As a consequence of the results contained herein, and a reexamination of Strong's work, the utility of the Strong Blank may be defined more precisely as indicative of an individual's continuance in an occupation.*

Routine clerical workers with high scores on the Strong Blank may have been attracted to their work as a result of those interests. However, for routine clerical jobs, possession of interests similar to those of successful office workers, as has been seen, was not an important factor contributing to job satisfaction. A whole complex of factors, in and away from the job, of which vocational interest is only one may have combined to produce satisfaction. Thus while clerical work may have attracted individuals who had high clerical scores on the Strong Blank, it did not insure that job satisfaction would follow.

Another possibility must be taken into account. Some of the clerks with high scores on the Strong clerical scale may have had primary vocational interests in other vocational areas. The presence of competing primary interests may have caused a diminution in the degree of job satisfaction experienced by clerical workers who had high Strong clerical interest scores. These competing interests may not have found fulfillment in what the clerks were working at and may have caused a lowering of job satisfaction scores. Research is necessary to ascertain whether two competing and apparently equally strong interests prevent the gaining of satisfaction from the pursuit of one of these occupations.

Increasing importance has been assigned to vocational interest patterns rather than scores on single scales. Strong and Darley suggested that guidance should not rely upon single interest scale scores, but should have as its basis scores earned within separate interest patterns. Possibly if the Strong Vocational Interest Blanks had been scored according to interest patterns, there might have been more correspondence with job satisfaction scores. This possibility should be investigated for its practical import to guidance and for determining the comparative value of single scales as compared with interest patterns.

RORSCHACH ADJUSTMENT

The evaluation of the clerks' Rorschach responses using the Davidson signs resulted in mean scores that permit conflicting interpretations. When compared with Davidson's original group (10), there is a wide difference in the number of adjustment signs, while the groups that Margulies (35) tested do not appear to be much different. The significance of differences between the groups was not tested because both Margulies' and Davidson's groups were composed of children and because only 16 of the 17 Davidson signs were utilized in evaluating the Rorschach responses of the clerks.

The Munroe Inspection Technique seems to demonstrate clearly that the clerks exhibited more Rorschach signs of maladjustment when compared with college students, artists, scientists, and technicians. The Munroe Technique presumably affords a fairer appraisal than the Davidson List inasmuch as the compared groups were composed of adults. When so considered, the results agree with those of Dodge (11) who found that the successful routine clerks considered themselves more ill at ease in social contacts, less willing to accept responsibility, and more dependent upon others. According to standard Rorschach interpretation these traits would be among those contributing to scores for maladjustment. While the clerks had many Rorschach signs of maladjustment, it is possible that these traits were ones that aided them in adapting to their work. Without these traits, it is conceivable that these routine clerks might have experienced difficulty in accepting the conditions of their work. In his study of 142 psychoneurotic ex-soldiers, Pratt revealed that there was a definite tendency to enter clerical work after having failed to adjust to other types of employment (44). Clerical work was also reported to have helped relieve their symptoms (44). Another suggested possibility is that clerical work is a refuge of individuals with certain characteristics. These deductions would apply to routine clerical workers and would not be descriptive of administrative clerks, or clerks whose jobs entail extensive public contact.

JOB SATISFACTION AND RORSCHACH ADJUSTMENT

There was no apparent relationship between the degree of job satisfaction and scores on either of the objective son List is not as extensive as the Munroe Technique nor are there any such fine distinctions in the weightings assigned to differences in quality and quantity of Rorschach responses.

PREDICTION OF CLERICAL INTERESTS FROM MUNROE INSPECTION LIST CATEGORIES

The multiple correlation was .72 between clerical interest scores on the Strong Blank and the following five Rorschach categories on the Munroe Inspection Technique: CF:FC, FM:M, FC, M, and Popular. This represented the correlation between the Strong clerical interest scores and the best estimates of these Strong scores made from a linear combination of the aforementioned five Munroe List categories. The multiple correlation coefficient was .70 after correction for shrinkage. Knowing the entries made for these five Munroe List categories, it would be possible with fair accuracy to predict Strong clerical interest scores of routine clerks. Forty-nine percent of the variance of the vocational interest scores of routine clerks could be attributed to the joint action of these five Munroe List categories. The remaining 51 percent of what made the routine clerks differ in vocational interest scores could be attributed to factors other than these five Munroe List categories.

VOCATIONAL INTERESTS AND RORSCHACH RESPONSES: THEORETICAL IMPLICATIONS

Routine clerks who showed interests similar to those of successful office workers were more likely than others to reveal what Rorschach examiners would call severe personality disturbances. The higher the scores on the clerical scale of the Strong Blank, the more numerous were the signs of personality maladjustment in their Rorschach responses.

According to standard Rorschach interpretation, routine clerks who had interests most similar to those of successful office workers were prone:

1. To impulsive emotional reactions to stimuli (the interpretation of the predominance of indefinite form-color [CF] over definite form color [FC] responses). These clerks had slight conscious control over their emotional reactions to stimuli. Their activities seemed largely determined and provoked by immediate environmental stimuli.
2. To show an inadequate degree of emotional adjustment to outer reality (the interpretation of an infrequent use of definite form-color [FC] responses). These clerks, apparently, were not able to make adequate social contacts with their environment.
3. To be emotionally immature (the interpretation of the predominance of animal action [FM] over human action [M] responses). They were prone to be dependent individuals whose inner strivings were mainly on an immature level.
4. To be unable to accept their own ideas and creations (the interpretation of the infrequent use of human action [M] responses). They did not seem to take their own outlook and ideas seriously.
5. To show a slight incapacity to think along conventional lines (the interpretation of the infrequent use of popular responses). These subjects appeared to show a slight inability to indulge in thinking common to most people.

The section on Rorschach interpretations has been offered with much hesitation, in view of the many unsubstantiated interpretations and varying meanings assigned to the Rorschach scoring symbols by some Rorschach examiners. It has been included only to describe the Rorschach characteristics of routine clerks and the associations of these Rorschach responses with vocational interests and job satisfaction in terms of the more commonly accepted interpretations of Rorschach examiners.

FURTHER RESEARCH SUGGESTIONS

The relation between the lack of primary interest patterns and job satisfaction should be ascertained. The relation between the lack of primary interest patterns and personality traits such as those the Munroe Inspection List was intended to assess should be investigated. The three nonoccupational scales of the Strong Vocational Interest Blank should also be related to measures of job satisfaction and personality.

It would be worthwhile to score a group of life insurance salesmen on their own occupational scale and upon scales such as those for office workers or mathematicians, and to relate these scores to the Rorschach categories of the Munroe Inspection List. Mathematicians and life insurance salesmen seem to represent opposite extremes in their social predilections. Will such tendencies be reflected in the Rorschach Test results and the degree of association of Munroe Inspection List categories to interest scale scores?

A developmental or cross-sectional study of young men at ages 17, 19, 21 and 23 using the Rorschach Test and the Strong Vocational Interest Blank would be very worthwhile. Changes in scores on the appropriate interest scales in relation to changes in the several categories of the Munroe Inspection List would help in clarifying the relation of personality patterns to vocational interests.

Another suggestion would envisage the testing with the Rorschach Test and the Strong Blank of students at a business school, clerks who had been on the job between six months and one year, and clerks who had been on the job 5 years or more. Would these groups show progressively higher mean scores on the Strong clerical scale? It would be worthwhile to ascertain whether certain Rorschach factors, apparently associated with the possession of interests similar to those of successful office workers, became more dominant in clerks with greater clerical job experience.

CHAPTER VI

SUMMARY AND CONCLUSIONS

Problem

THE aim of this study was to investigate the vocational interests, Rorschach responses, and job satisfaction of a group of 100 men employed in routine clerical work by the Federal government. The interrelationships among such factors as job satisfaction, vocational interests, and Rorschach signs of maladjustment were sought. Those Rorschach responses which were related significantly to job satisfaction and vocational interests were isolated. The subjects were chosen so that they fulfilled the following conditions: (*a*) permanence and stability of employment; (*b*) similarity of occupation; (*c*) presumably similar general material working conditions; (*d*) restricted salary range; (*e*) employment in a vocation for which a scoring scale had been established on the Strong Vocational Interest Blank.

It was desired, through this study, to ascertain if the following hypotheses were tenable:

1. The job satisfaction of routine office clerks bears a significant positive relationship to the possession of interests similar to those of successful office workers as measured by the Strong Vocational Interest Blank.

2. The job satisfaction of routine clerks is not associated with the number of Rorschach signs of maladjustment.

3. The work dissatisfaction of routine clerks who possess the interests of successful office workers is not related to the number of Rorschach signs of maladjustment.

4. Certain Rorschach responses shown in the Rorschach records of these clerks are significantly associated with the possession of vocational interests held by successful office workers.

Procedure

The 100 routine clerks completed the job satisfaction blank, the Strong Vocational Interest Blank, and the Rorschach Test. The administration of the tests to the subjects took place in groups of four or five. Anonymity was guaranteed to all. Standard procedures were followed with the job satisfaction blank and the Strong Vocational Interest Blank. As for the Rorschach Test, the clerks after being given sets of cards, wrote their responses in special booklets, and separate individual inquiries were conducted for all the protocols.

The Strong Blank was scored with the office worker scale devised by Strong. Scores were computed for the job satisfaction blank. The Rorschach protocols were evaluated by two techniques, the Munroe Inspection Technique and the Davidson List of Signs. Numerical scores were obtained for each list, indicative of the degree of Rorschach maladjustment and adjustment. The relationships between each of the separate measures were obtained through product-moment correlations, biserial correlations, partial correlations, and analysis of variance and chi-square methods. In addition, several of the categories in the Munroe Inspection Technique were related to job satisfaction and to Strong clerical interest scores. Correlations among five of the categories on the Munroe Inspection Technique were computed by product moment and tetrachoric methods. Finally, a multiple correlation coefficient was calculated for the prediction of the Strong clerical interest scores of routine office clerks from five categories of the Munroe Inspection List.

Findings

The *first hypothesis* was not supported by the results of this investigation. The job satisfaction of routine clerks appeared to have no significant relationship to the scores and letter ratings earned on the clerical scale of the Strong Vocational Interest Blank. The differences between the criterion group used by Strong in standardizing the clerical interest scale and the routine clerks were pointed out. It was suggested that if a vocational interest scale based upon routine clerical workers had been utilized, greater association between the degree of job satisfaction and the possession of the corresponding vocational interests might have been demonstrated. Moreover, other reasons were advanced to support the concept that satisfaction with routine clerical work might not be associated with the possession of the interests of successful office workers. The nature of routine clerical work and the relative lack of demand that it makes on the skills and talents of the incumbents were underscored. A close relationship between job satisfaction and vocational interests was posited for professional and skilled mechanical workers in view of the relative satisfyingness of achieving something creative or difficult. Little such opportunity exists for routine clerks. The presence of competing vocational interests may have caused a diminution in the association between job satisfaction and vocational interests of routine clerks but was not investigated. Vocational interest pattern scores, embracing a family of related occupations, may have shown greater relationship to job satisfaction than single interest scale scores, but this possibility also was not investigated. Attention was directed to the fact that while vocational interests may have attracted some of these routine clerks to their work, other factors may be more important in the occupational choice and job satisfaction of clerical workers than vocational interests.

The *second hypothesis* was confirmed: there was no relationship between job satisfaction and the number of Rorschach signs of maladjustment for routine clerks. Routine clerks may be satisfied with their work despite many Rorschach signs of maladjustment. Apparently whatever is revealed by many Rorschach signs of maladjustment did not prevent routine clerks from deriving satisfaction from their work. The validity of the Rorschach Test was questioned because it failed to distinguish between the satisfied on the one hand and the indifferent and dissatisfied on the other hand. An equally credible alternative was that the reasons for work satisfaction or dissatisfaction in office clerks should not be sought in the psychological context, i.e., in responses to the Strong Vocational Interest Blank and the Rorschach Test, but rather in the social situation, as suggested by recent intensive investigations.

The *third hypothesis* was sustained: the job dissatisfaction or indifference of routine clerks who possessed the interests of successful office workers was not associated with a greater number of Rorschach signs of maladjustment. The presence of many Rorschach signs of maladjustment did not detract from or add to the job satisfaction of routine clerical workers who have the interests of successful office workers.

The *fourth hypothesis* was supported by the results. There were five Rorschach response categories significantly associated with the possession of vocational interests by routine clerks similar to those of successful office workers.

These five Rorschach signs of maladjustment that were correlated significantly with the possession of vocational interests similar to those of successful office workers included: (*a*) the use of indefinite form-color (CF) responses in preference to definite form fused with color (FC) responses; (*b*) a slight tendency to use an insufficient number of definite form-color (FC) responses; (*c*) a proneness to use predominantly more animal action (FM) than human action (M) responses; (*d*) relatively few human action (M) responses; (*e*) slightly but significantly infrequent use of popular responses.

The Rorschach characteristics of routine clerks appeared to be quite similar to those of the group with which they were contrasted except for the clerks' greater use of animal action (FM) responses. This characteristic was indicated as a Rorschach sign of maladjustment that the clerical group as a whole manifested.

The two measures of Rorschach adjustment, the Davidson List of Adjustment Signs and the Munroe Inspection Technique, were shown to be highly correlated. It was concluded that they were both measuring a similar pattern or similar patterns. The extent to which the two lists arrive at different evaluations of adjustment is a function of their two major differences: the Davidson List is not as extensive, and does not include such fine distinctions in quality and quantity of the Rorschach responses.

Intercorrelations were obtained between several of the Rorschach categories on the Munroe Inspection List, and a multiple correlation coefficient of .70 was computed for the prediction of Strong clerical interest scores from these several categories.

Further supplementary areas of research were delineated and several hypotheses were discussed which still require substantiation.

Appendix A

JOB SATISFACTION BLANK
(After Hoppock, 43)

NUMBER ____________________

You are asked to help in a scientific study by answering the questions in this blank. Neither your employer nor any of your associates will be allowed to see your answers. Your replies will be added to those of many other people, and only the group totals will be published. Your answers will be worthless unless they are perfectly frank and truthful.

Choose the ONE of the following statements which best tells how well you like your job. Place a mark (X) in front of that statement.

1. ____I hate it.
2. ____I dislike it very much.
3. ____I don't like it.
4. ____I am indifferent to it.
5. ____I like it.
6. ____I am enthusiastic about it.
7. ____I like it more than I could any other job.

Mark one of the following to show how you think you compare with other people.

7. ____No one likes his job better than I like mine.
6. ____I like my job much better than most people like theirs.
5. ____I like my job better than most people like theirs.
4. ____I like my job about as well as most people like theirs.
3. ____I dislike my job more than most people like theirs.
2. ____I dislike my job much more than most people dislike theirs.
1. ____No one dislikes his job more than I dislike mine.

Mark the one of the following which best tells how you feel about changing your job.

1. ____I would quit the job at once if I could get anything else to do.
2. ____I would take almost any other job in which I could earn as much as I am earning now.
3. ____I would like to change both my job and my occupation.
4. ____I would like to exchange my present job for another job in the same line of work.
5. ____I am not eager to change my job, but would do so if I could get a better job.
6. ____I cannot think of any jobs for which I would exchange mine.
7. ____I would not exchange my job for any others.

Mark one of the following to show HOW MUCH OF THE TIME you feel satisfied with your job.

7. ____All of the time.
6. ____Most of the time.
5. ____A good deal of the time.
4. ____About half of the time.
3. ____Occasionally.
2. ____Seldom.
1. ____Never.

Date ____________________

APPENDIX B

ARBITRARY WEIGHTS, AND ENTRIES CONSIDERED CORRECT FOR BISERIAL AND TETRACHORIC CORRELATIONS
(MUNROE INSPECTION LIST CATEGORIES)

PRODUCT MOMENT CORRELATION WEIGHTS

Color:Movement		FM:M		M		CF:FC		FC	
Entries	Weights	Entries	Weights	Entries	Weights	Entries	Weights	Entries	Weights
− − −	1	+ + +	1	+ + +	1	+ + +	1	− −	6
− −	2	+ +	2	+ +	2	+ +	2	−	5
−	3	+	3	+	3	+	3	o	4
o	4	o	4	o	4	o	4		
+	5	−	5	−	5	−	5		
+ +	6	− −	6	− −	6				
+ + +	7								

ENTRIES CONSIDERED CORRECT FOR BISERIAL AND TETRACHORIC CORRELATIONS

Category	Entries Considered Correct		
Color:Movement	− − −	− −	−
CF:FC		−	o
FC		− −	−
FM:M	− −	−	o
M		− −	−
Total Color		− −	−
Total Movement		+ +	+
FK, Fc		− −	−
F%		− −	−
Range	− − −	− −	−
Popular	− − −	− −	−
Dd	+ + +	+ +	+
W			+

BIBLIOGRAPHY

1. Beck, Samuel J. *Rorschach's test.* 2 vols. New York: 1944-45, Grune and Stratton.
2. Beckman, R. O. A new scale for gauging occupational rank. *Personnel J.*, 1934-35, **13**, 225-233.
3. Berdie, Ralph F. The prediction of college achievement and satisfaction. *J. appl. Psychol.*, 1941, **32**, 197-204.
4. Berdie, Ralph F. Factors related to vocational interests. *Psychol. Bull.*, 1944, **41**, 137-155.
5. Bordin, Edward S. A theory of vocational interests as dynamic phenomena. *Educ. psychol. Measmt,* 1943, **3**, 49-65.
6. Brussel, James A., and Hitch, Kenneth S. *An introduction to Rorschach Psychodiagnostics in military and civilian psychiatry* and
   Piotrowski, Zygmunt A. *A Rorschach compendium.* Utica: 1947, State Hospital Press.
7. Carter, Harold D. The development of vocational attitudes. *J. consult. Psychol.*, 1940, **4**, 185-191.
8. Carter, Harold D. Vocational interests and job orientation. *Appl. Psychol. Monogr.* No. 2. Stanford: Stanford University Press, 1944.
9. Darley, John G. *Clinical aspects and interpretation of the Strong Vocational Interest Blank.* New York: Psychological Corp., 1941.
10. Davidson, Helen H. *Personality and economic background.* New York: Kings Crown Press, 1943.
11. Dodge, Arthur F. What are the personality traits of the successful clerical worker? *J. appl. Psychol.*, 1940, **24**, 576-586.
12. Dodge, Arthur F. Characteristics of good clerks. *Personnel J.*, 1942, **20**, 324-340.
13. Fairchild, M. Skill specialization, a study in the metal trades; Part I. The nature and measurement of skill; Part II. The significance of skill. *Personnel J.*, 1930, **9**, 28-71, 128-175.
14. Fisher, V. E., and Hanna, Joseph V. *The dissatisfied worker.* New York: Macmillan, 1931.
15. Ford, Mary. *The application of the Rorschach Test to young children.* Minneapolis: Univ. Minn. Press, 1946.
16. Ghiselli, E. F. and Brown, Clarence W. *Personnel and industrial psychology.* New York: McGraw-Hill, 1948.
17. Habbe, S. Job attitudes of life insurance agents. *J. appl. Psychol.*, 1947, **31**, 111-128.
18. Hand, T. Hoppock, R. and Zlatchin, P. Job satisfaction, researches of 1944 and 1945. *Occupations*, 1948, **26**, 425-431.
19. Harrower-Erickson, M. R. and Steiner, M. E. *Large scale Rorschach techniques.* Springfield, Ill.: Thomas, 1945.
20. Hertz, Marguerite R. The reliability of the Rorschach Ink Blot Test. *J. appl. Psychol.*, 1934, **18**, 461-477.
21. Hertz, Marguerite R. The Rorschach Ink Blot Test: historical summary. *Psychol. Bull.*, 1935, **32**, 33-66.
22. Hertz, Marguerite R. Personality patterns in adolescence as portrayed by the Rorschach Ink Blot method: I. The movement factors. *J. gen. Psychol.*, 1942, **27**, 119-188.
23. Hertz, Marguerite R. The validity of the Rorschach group method. *Psychol. Bull.*, 1942, **39**, 514-515.
24. Hertzman, Max. A comparison of the individual and group Rorschach tests. *Rorschach Res. Exch.*, 1942, **6**, 89-108.
25. Hoppock, Robert. *Job satisfaction.* New York: Harpers, 1935.
26. Horst, Paul. *Prediction of adjustment.* New York: Social Science Research Council, Bull. #48, 1941.
27. Kerr, M. The Rorschach Test applied to children. *Brit. J. Psychol.*, 1934, **25**, 170-185.
28. Klopfer, Bruno and Kelley, Douglas. *The Rorschach technique.* Yonkers, N.Y.: World Book, 1942.
29. Kornhauser, Arthur. Psychological studies of employee attitudes. *J. consult. Psychol.* 1944, **8**, 127-143.
30. Kuhlen, Raymond G. Age differences in personality: adult years. *Psychol. Bull.*, 1945, **42**, 333-358.
31. Lane, Barbara M. A validation test of the Rorschach movement interpretation. *Amer. J. Orthopsychiat.*, 1948, **18**, 292-296.
32. Levine, Kate N., Grassi, Joseph R. and Gerson, Martin J. Hypnotically induced mood changes in the verbal and graphic Rorschach: A case study Part II. The response records. *Rorschach Res. Exch.*, 1944, **8**, 104-124.
33. Lindner, R. M. A further contribution to the group Rorschach. *Rorschach Res. Exch.*, 1943, **7**, 7-15.
34. Lorimer, Margaret. An appraisal of vocational guidance. *J. higher Educ.*, 1944, **15**, 260-267.
35. Margulies, Helen. Rorschach responses of successful and unsuccessful students. *Arch. Psychol.*, New York, 1942, No. 271.
36. Munroe, Ruth L. An experiment in large scale testing by a modification of the Rorschach method. *J. Psychol.*, 1942, **13**, 229-263.

37. Munroe, Ruth L. The inspection technique, a method of rapid evaluation of the Rorschach protocol. *Rorschach Res. Exch.*, 1944, **8**, 46-70.
38. Munroe, Ruth L. Objective methods and the Rorschach blots. *Rorschach Res. Exch.*, 1945, **9**, 59-73.
39. Munroe, Ruth L. Rorschach findings on college students showing different constellations of subscores on the ACE. *J. consult. Psychol.*, 1946, **10**, 301-316.
40. Munroe, Ruth L. Prediction of the adjustment and academic performance of college students. *Appl. Psychol. Monogr.* No. 7. Stanford: Stanford Univ. Press, 1945.
41. Nahm, Helen. Satisfaction with nursing. *J. appl. Psychol.*, 1948, **32**, 335-343.
42. Paterson, D. G. and Stone, H. C. Dissatisfaction with life work among adult workers. *Occupations*, 1942, **21**, 219-221.
43. Pond, Millicent and Bills, Marion A. Intelligence and clerical jobs: two studies of relation of test score to job held. *Personnel J.*, 1933, **12**, 189-196.
44. Pratt, Dallas. Reemployment of the psychoneurotic ex-soldier. *Psychiatry*, 1945, **8**, 3-8.
45. Roe, Anne. Alcohol and creative work: painters. *Quart. J. Studies in Alcohol*, 1946, **6**, 415-467.
46. Roe, Anne. Painting and personality. *Rorschach Res. Exch.*, 1946, **10**, 86-100.
47. Roe, Anne. A Rorschach study of a group of scientists and technicians. *J. consult. Psychol.*, 1946, **10**, 317-327.
48. Roethlisberger, F. J. and Dickson, W. J. *Management and the worker.* Cambridge: Harvard Univ. Press, 1939.
49. Rorschach, Hermann. *Psychodiagnostik.* Berne, Switzerland: Hans Huber, 1942.
50. Sarbin, T. R. and Anderson, H. C. A preliminary study of the relation of measured interest patterns and occupational dissatisfaction. *Educ. psychol. Measmt.*, 1942, **2**, 23-26.
51. Shaffer, Laurance F. *Psychology of adjustment.* New York: Houghton-Mifflin, 1936.
52. Strong, E. K., Jr. *Vocational interests of men and women.* Stanford: Stanford Univ. Press, 1943.
53. Super, Donald E. Occupational level and job satisfaction. *J. appl. Psychol.*, 1939, **23**, 547-564.
54. Super, Donald E. Strong's vocational interests of men and women; special review. *Psychol. Bull.*, 1945, **42**, 359-370.
55. Swift, Joan W. Reliability of Rorschach scoring categories with pre-school children. *Child Develpm.*, 1944, **15**, 207-216.
56. Thorndike, Robert L. (Ed.) *Research problems and techniques, report No. 3.* Army Air Forces Aviation Psychology Program Research Reports, 1946.
57. Uhrbrock, R. Attitudes of 4,430 employees. *J. soc. Psychol.*, 1934, **5**, 365-377.
58. Van Dusen, A. C. Permanence of vocational interests. *J. educ. Psychol.*, 1940, **31**, 401-424.
59. Vernon, P. E. The Rorschach Ink-Blot Test. *Brit. J. med. Psychol.*, 1933, **13**, 89-114, 179-200, 271, 291.
60. Vernon, P. E. Recent work in the Rorschach Test. *J. ment. Sci.*, 1935, **81**, 894-920.
61. Watson, Goodwin. Morale during unemployment. In Goodwin Watson (Ed.). *Civilian morale*, New York: Houghton-Mifflin, 1942, Pp. 273-348.
62. Wesman, Alexander. *Personnel Classification Test manual.* New York: Psychological Corp., 1947.
63. Woods, Walter A. Employee attitudes and their relation to morale. *J. appl. Psychol.*, 1944, **28**, 285-301.
64. Wrenn, C. Gilbert. Vocational satisfaction of Stanford graduates. *Personnel J.*, 1934, **13**, 21-24.
65. Young, Kimball. *Personality and problems of adjustment.* New York: Croft, 1940.
66. Zubin, Joseph. A psychometric approach to the evaluation of the Rorschach Test. *Psychiatry*, 1941, **4**, 547-566.

VOLUME 64
NUMBER 4

WHOLE NO. 310
1950

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology* with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

Symbol Elaboration Test (S.E.T.):

The Reliability and Validity of a New Projective Technique

By
JOHANNA KROUT
Chicago Psychological Institute
Chicago 5, Ill.

Accepted for publication Sept. 20, 1949

Price $2.00

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION
1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

COPYRIGHT, 1950, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION

ACKNOWLEDGMENT

THE writer wishes to record her gratitude to Drs. William E. Henry, Robert J. Havighurst, and Carolyn Tryon, of the Committee on Human Development of the University of Chicago, who, throughout the preparation of this monograph, offered patient criticism and guidance, often at the expense of personal convenience. She also feels indebted to Dr. Donald W. MacKinnon of the University of California, and Dr. Herbert S. Conrad of the United States Office of Education, for their valuable editorial suggestions.

In addition, she wishes to thank the following colleagues whose cooperation gave this study greater objectivity and saved time in preparation: Drs. Theodore J. Dulin, Maurice H. Krout, Carson Macguire, Jules Masserman, and Winifred Shannon; Misses Norma Levenfeld, Hilde Richard, and Marian Sobel; and Messers Harold Goldberg and Arnold Palevsky. Thanks are also due to the Committee on Human Development of the University of Chicago, the Chicago Psychological Institute, and the United States Indian Bureau for providing subjects for this study.

JOHANNA KROUT

TABLE OF CONTENTS

I. PROBLEM 1
BACKGROUND STUDIES 1
BASIC CONCEPTS 2
EXPERIMENTAL DESIGN 3
SELECTION OF STIMULUS PATTERNS 4
The semi-circle pattern 4
The counter-opposed semi-circle pattern 5
The semi-circle within a larger semi-circle pattern 6
The single straight line pattern 7
The crossline pattern 7
The semi-circle and single line pattern 9
The lines and semi-circle pattern 9
The incomplete "U" pattern 9
The diffuse mass pattern 10
The inverted "V" pattern 10
The jagged line pattern 11
II. PROCEDURE 12
PRESENTATION OF STIMULI 12
ANALYSIS OF STIMULUS MATERIAL 13
VALIDATION DATA FOR EACH SUBJECT GROUP 17
Normal children 18
Normal adults 18
Children with behavior problems 19
Adults in psychotherapy 19
Indian group 19
SYSTEM FOR VALIDATION-RATING 20
CHECKS FOR RELIABILITY 22
Normal children group 22
Adult and problem children groups 22

Indian group 24

Method of analysis 24

III. SAMPLE ANALYSES 25

Source of material 25

Case A: White boy, age 13 25

Case B. White female, age 24 33

IV. RESULTS 51

Analysis of total results 51

Sex factors 55

Factor of emotional stability 57

Factor of age 60

Stimulus sequence as a factor 63

V. CONCLUSIONS 65

General implications of this study 65

Use of the symbols as basis for projective technique 65

CHAPTER I

PROBLEM

A. BACKGROUND STUDIES

THE first experimental attack on the problem of symbolism is credited by Sigmund Freud (3-a) to Scherner. In his attempt to establish the causes underlying dream symbols, Scherner investigated the relationship between various stimuli and the dream-work. From his experiments he concluded that dream symbols are the product of physiological tensions.

Freud (3-b) felt that Scherner had not gone far enough into the problem. In working out the patterns of the dream process, Freud had observed that certain forms repeated themselves, with variations, in different subjects. These forms seemed to have the same symbolic significance; although how the patient felt about them was indicated by other aspects of the dream-work, and varied from individual to individual. Stekel (21), Gutheil (8), and others later confirmed Freud's findings.

Psychoanalysts as a group have contributed much of the interest in, and most of the discussion about, universal symbols in recent times. They differ in their viewpoints, however. Some analysts such as Jung (12), White (23), and Barran (2), impressed by parallels between primitive myths and the dream-work of their patients, have concluded that there must be an "archaic unconscious," common to all people, which consists of the inherited experiences of earliest man. Horney (10) took exception to this. She has held that cultural factors and variations in individual experience make it impossible to generalize on symbolic meanings to the extent of assuming the existence of universal symbols.

Others, like Ferenczi (5), and Symons (22), seem to have formed the clinical impression that universal symbols do exist. However, they have not found it necessary to resort to constructs like the "archaic unconscious" to explain the similarity between private symbol-systems, such as are found in dreams, and primitive myths. Neither have they found it necessary to ignore such a similarity because of chronologically secondary, though clinically important, indivdual differences. They base their theory on the belief that the process of thought is similar in all people.

Some non-analysts will not accept the concept of universal symbolism, but not so much because they are opposed to the concept as such. What they oppose is the methods used to establish universal symbolism. Havelock Ellis has fought the subjective bias in psychoanalysis which, he said, was at the basis of the belief in universal symbolism. He thought that the analysts, unwittingly or by design, have communicated to patients their own conceptions of what certain symbols mean. The possibility that this might be true raises certain questions as to the reliability, not to say validity, of various types of clinical findings.

Anthropologists have been concerned with the problem of universality in human experience. The present tendency is to emphasize cultural differences. However, some workers have been even more impressed by apparently basic similarities. For example, Roheim (18) established that there are certain symbols which appear among all the Australian tribes he had studied. Hallowell (9) has found it possible to analyze an Indian's dream in terms of Freudian symbols.

Various concepts growing out of psychoanalysis have been checked experimentally by Sears (19) and others, but the concept of symbolism as such has never been studied by experimental methods. Interest in projective techniques recently has drawn attention to the problem anew. The Rorschach test in particular, in its "qualitative" or symbolic-content aspects, carries the implication that universal symbols can be assumed to exist. Lindner (14) wrote about diagnostic Rorschach signs in terms of this implication. However, Lindner's observations also were in the nature of clinical findings rather than experimental results.

Among psychologists, Franck (7), assuming the correctness of universal symbolism, constructed a test of esthetic preference based on "masculine" and "feminine" figures. Similarly Buck (4) in constructing his House—Tree—Person (H-T-P) test, has posited a certain universalism of response as he generalized on the cognitive aspects of his symbols.

Alschuler and Hattwick (1) have made a comprehensive study of symbolism in the paintings of young children. The method of these authors was completely inductive. When they compared the free drawings of children, they found evidence for assuming definite uniformities in the images of their subjects. Similar conclusions were reached by Immergluck and Cohen (11) who instructed their subjects to react to emotionally charged words (e.g., mother, father, love, war) by chance designs produced on paper. They found striking similarities in rhythmic patterns produced in association with given word stimuli. This study is especially interesting because it was not the purpose of these investigators to establish the possibility of symbolism in any form.

In summarizing the theoretical material in the literature pertaining to cultural symbolism, we may conclude that psychoanalysts in general have accepted the existence of universal symbols. They have found the concept useful in dream analysis and the interpretation of symptoms in the course of therapy. Anthropologists have applied the same concept to the analysis of myths of different peoples. However, because psychoanalysts and anthropologists have used either clinical or deductive methods for the most part, psychologists as a rule have declined to accept their conclusions. Recent psychological studies, such as those of Franck, Buck, Alschuler and Hattwick, and Immergluck and Cohen, have either assumed or inferred symbolic values, but did not aim to establish them.

This, substantially, was the state of the literature when the present study was begun in the winter of 1946.

B. Basic Concepts

This study is concerned with reactions to certain visual stimuli. These stimuli have two characteristics in common. First, they are line figures which might be called *nonsense patterns,* because they are not pictorial representations of any obvious object, person or human detail; second, they have been presumed to possess a universally symbolic value. Because neither these nor any other line-patterns have been accepted as universal symbols by psychologists, this study raises the general question of whether universal symbols may be said to exist at all.

Clarity of discussion demands a definition of terms. Before details are given as to how this study was designed, we must make clear what we mean here by universal symbols. There are three points to consider: 1. What is a symbol? 2. What implications does the concept of universality carry in this study? 3. How can supposed nonsense figures actually be universal symbols?

First, the term *symbol* has been given a variety of meanings by various people, depending upon the application involved. There are as many definitions for the word "symbol" as there are authors writing on the subject, but the layman's meaning of "something that stands for something else" is preserved in all of them. For our purposes, the definition given by C. W. Morris (16) is most acceptable in that it emphasizes the stimulus-value of a symbol. He defines the symbol as a "given (or experienced) substitute-stimulus leading to a reinstatement of the original stimulus in a form observable only from the self-inclusive point of view." This definition implies that all symbols affect behavior by first evoking previous experience. It also implies that the image of the original experience is always peculiar to the individual.

A symbol which is effective for more than one person is based upon experience that has been shared by all for whom the symbol is effective. However, since any experience is uniquely personal (in the psychobiological sense), the reaction to any symbol is also uniquely personal. The Morris (16) definition holds for all symbols, whether verbal or non-verbal. Just how it applies may most easily be seen in a verbal sample. When an English-speaking person is given the stimulus word, "Mother," he responds both in terms of: (*a*) the concept "Mother" as shared by all individuals (e.g., the quality of bearing children) and (*b*) the image of his own mother as it remains in his memory from his own experiences with her (e.g., enforcement of cleanliness, emphasis on religion, etc.). In other words his response contains elements which are found in the responses of others and elements which no one else's responses could contain.

Second, to assume that a symbol possesses a general (as distinquished from a uniquely personal) value, is to lay claim to two points: One, that there are experiences which are general, so that responses to certain symbols include reinstatements of original stimuli which are general for all people; and two, that these experiences may be symbolized by a form which is capable of reinstating the original (personal) experience. That generalized cultural (or reactive) experiences do exist among all peoples has been established by numerous anthropologists. A recent study by Ford (6) has demonstrated universal behavior patterns in 25 preliterate groups. Some of the patterns for which he listed 100% universality are eating, talking, coitus, etc.

Perception of structural differences in human beings may be added as a fairly general experience among people. Thus Seligman (20) has made the point that similar biological characteristics of human individuals are responsible for the universal features of certain culture patterns. An unanswered question is whether the *form* of symbolization is generalized enough to include even various individuals in a given culture, let alone all people everywhere. This is an issue on which the present study may throw some light but on which it will not seek to offer final proof.

The final question that arises here is how seeming nonsense patterns could have a generalized symbolic value. The patterns involved are obviously abstract, i.e., remote from concrete human experience. When individuals respond to them nonetheless, it is surely not safe to assume that their response involves a conscious recognition of the patterns as symbols. It is possible, however, to assume that these abstract patterns are substitute (symbolic) stimuli which are recognized, though not on a level of clear understanding, as having a vague relationship to the original stimulus. Such perception, it must be emphasized, is not on a level of conscious awareness. In other words, the stimuli here operate without becoming understandable in verbal terms. The term "unconscious" is sometimes used to describe responses to the stimuli in question; but because of semantic difficulties, this term is here avoided, and the historically more respectable term "subliminal" (15) is used instead.

In summary, a generic or general symbol is defined as a substitute stimulus capable of reinstating a basic human experience which is part of the shared experience of many people in a given culture at least. Such a symbol may be so far removed from phenomenal reality as we know it that it can operate only as a subliminal cue. The response to it will be general only in the sense that the symbol will be an effective stimulus for all people sharing certain basic experiences. However, each person's response to such a "generalized symbol," as indeed to less common stimuli, must be assumed to include elements that are both unique and personal.

C. EXPERIMENTAL DESIGN

Eleven hypothetical symbols representing as many basic human experiences were included in this study. These symbols were designated as *stimulus patterns,* and were presented to the subjects as undifferentiated stimulus material.

The reactions of subjects to the stimulus patterns were obtained by the following method: The subjects were handed pencils and eleven sheets on each of which appeared one of the stimulus patterns. The directions asked for drawings of anything they pleased to present on each sheet, using the lines already on the sheet, if they could (details in Chapter II, section on *Presentation of Stimuli*). Next, an inquiry was conducted to establish what they had done and how they had felt about each of their drawings. In this way, not only the subjects' drawings, but their verbal associations with the stimulus material were obtained as well. In each case, the two kinds of reactions, the drawings and the verbal associations, were considered the *subject's response.*

In this study it was assumed that the subjects' responses covered areas of experience of which the stimulus-patterns were symbolic. In the light of this assumption it was expected that specific attitudes of the subjects growing out of these areas might be revealed in their responses. To establish or disprove this possibility, the interpretations of the subjects' responses were checked against material gathered on these same subjects by independent workers. The material used for validation was referred to as *validation data.*

In brief, then, the design of this study involved the assumption that, if a given stimulus possesses a generalized symbolic value, it should be possible to infer specific facts about a subject from his response to that stimulus. These facts would relate to the basic human experience symbolized, in terms of its meaning for the subject. Conversely, if it is possible to discover accurate and specific facts about a subject from his reactions to a stimulus with a hypothetically generalized symbolic value, the repetition of this procedure with many subjects should substantiate the hypothetical meaning of the stimulus.

D. Selection of Stimulus Patterns

Eleven basic experiences were selected as probably common to all human personalities in our culture, and possibly other cultures as well.[1] Visual forms were then adapted from various sources (See *Background Studies*) as symbolic of each area of experience. Thus eleven visual forms came to be used as stimulus-patterns in this study. Responses to the stimulus-patterns were considered to have been determined by the subjects' attitudes toward the "reinstatement" of the "original stimulus" (the basic experience) evoked by its supposed substitute-stimulus (the hypothetically symbolic stimulus pattern).

All of the stimulus-patterns were presented in as slightly structuralized form as possible,[2] so as to provide maximum opportunity for modification by the subject. The following is a list of the symbols selected and the sources from which they were adapted. It is considered beyond the scope of this study to establish why these particular symbols relate to the areas symbolized. However, theoretical reasons are suggested.

1. *The semi-circle pattern.* (Fig. 1). It seems to be a safe assumption that any human society provides its members with the experience of "femaleness." This implies only that every individual learns to differentiate between the two sexes. How a given individual interprets "femaleness" depends upon his personal experience with females. As stated in the discussion of

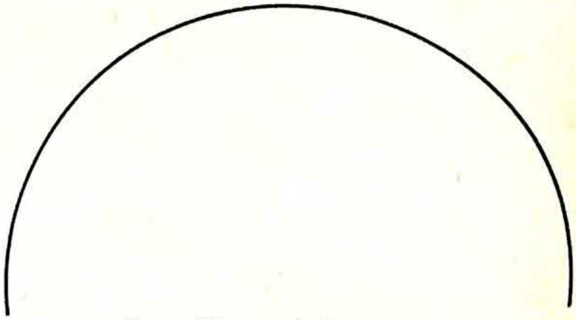


Fig. 1. The semi-circle pattern.

---

[1] Copyright of patterns symbolizing these experiences, as presented in this section, is reserved.

[2] The actual order of presentation was: (1) blank sheet, (2) semi-circle, (3) semi-circle and single line, (4) single straight line, (5) inverted "V," (6) diffuse mass, (7) crosslines, (8) counter-opposed semi-circle, (9) semi-circle within larger semi-circle, (10) jagged line, (11) incomplete "U," (12) lines and semi-circles.

"Basic Concepts," the image evoked by any generalization is necessarily peculiar and private to the subject. It is probably true that similarities between images of people from a given culture are greater than those between people from different cultures. The hypothesis to be tested by the first stimulus-pattern is that it is symbolic of "femaleness," or refers to aspects of "femaleness" which are fundamental to any such generalization.

The symbol for this basic experience originated with Freud (3-c, p. 378). It was his clinical observations that led him to assume that roundness is symbolically associated with the concept of "femaleness." Alschuler and Hattwick (1), in their study of children's drawings have found that both rounded forms and square ones were characteristic of children when they were behaving along lines of more "feminine" patterns. Also when their children reported negative feelings about their mothers, they tended to smear or otherwise destroy rounded forms in their drawings.

Franck (7) based her work on the concept that "feminine" design was composed of round forms. Similarly, Immergluck and Cohen (11), in classifying drawing patterns obtained in response to verbal stimuli, found unexpectedly that round lines tended to be associated with the concepts "mother" and "love."

In the light of such findings, the author has chosen the semi-circle, as a rounded form, to represent the most likely symbol for "femaleness." The drawing responses associated with this form were expected to reflect the subject's attitudes toward females, within the orbit of his or her personal experience.

2. *The counter-opposed semi-circle pattern.* (Fig. 2). One semi-circle has been assumed to mobilize attitudes toward the quality we designate as "femaleness." Going one step further, we could assume that this same concept could be utilized to test attitudes toward females generalized as "female society." For this reason, it seemed desirable to devise a field containing more than one semi-circle.

The pattern in question derives from a basic rationale which is thus similar to that of the first pattern, except for the aspect of pluralism which it is meant to convey. To avoid the implication of two breasts, which might refer to one woman, the two semi-circles were given a 90° tilt. In thus counter-opposing two symbols of the same independent significance, the pattern was presumably insured against perception in its singular, instead of in its plural, form.

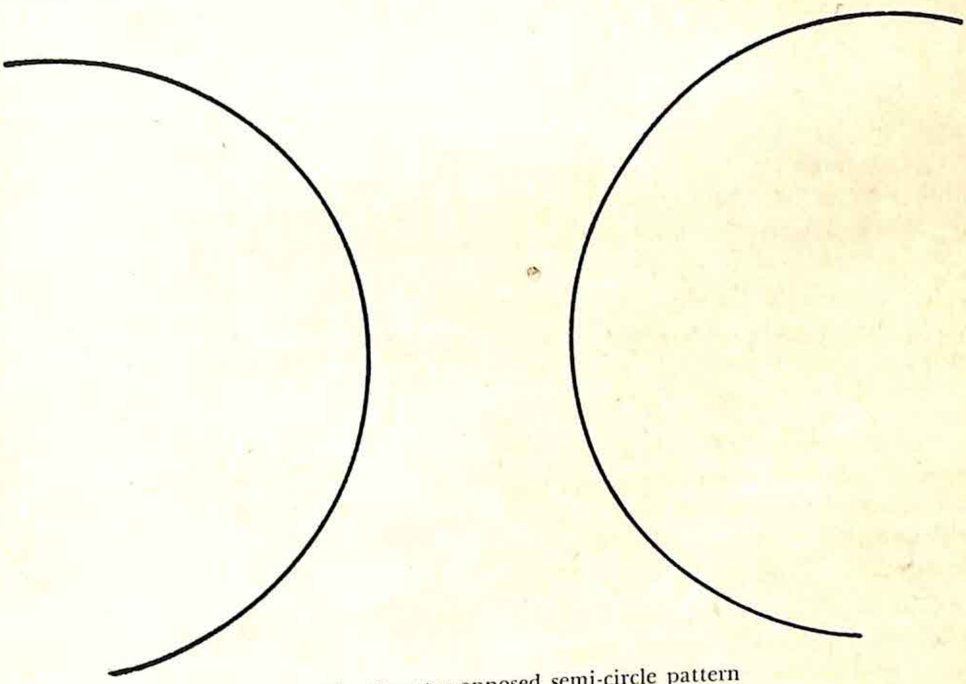


FIG. 2. The Counter-opposed semi-circle pattern

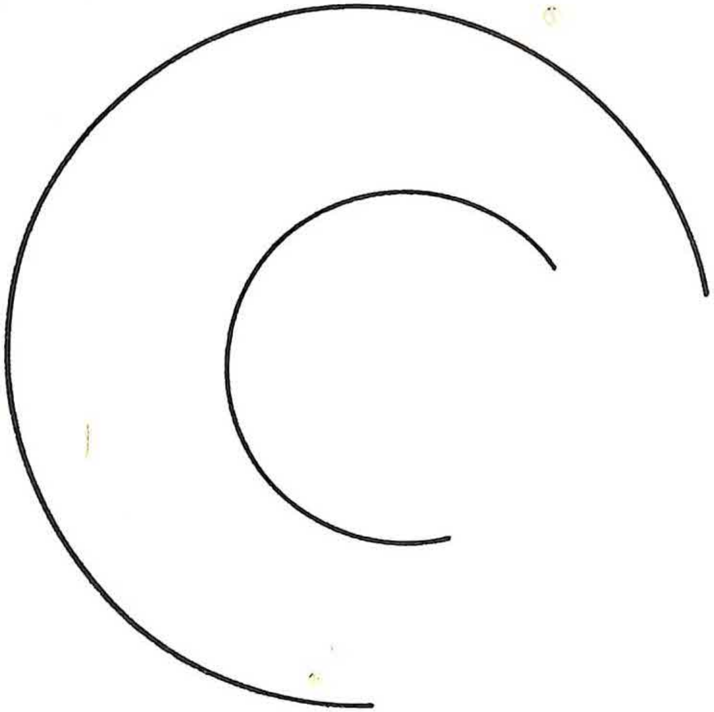


FIG. 3. The semi-circle within a larger semi-circle pattern.

If the assumption made is valid, then a female subject, identifying with one of the semi-circles, might reveal her attitude toward another woman; while a male subject, unable to find ground for identification in this figure, might show his attitude toward more than one woman.

3. *The semi-circle within a larger semi-circle pattern.* (Fig. 3). Adjustment to the increased responsibility which growing older engenders at every stage, may be considered a broadly human area of experience. The desire to regress, measured by the strength of dependence on the mother, and attitudes toward siblings, are all related to this area.

Roheim (18, p. 102) discusses what he calls the "concentric circle pattern" in relation to the fertility rites of Australian primitives. His treatment of the subject suggests that this might be a universal symbol for the "child and mother" area of experience. Freud (3, p. 375) of course made a good deal of this kind of symbolization. Alschuler and Hattwick (1) have found that rounded figures, or square ones, drawn with other geometrical forms placed inside, were characteristic of children's drawings shortly preceding the birth of younger siblings, and persisted for a time after their birth. These con-

FIG. 4. The single straight line pattern.

clusions support the Freudian hypothesis that forms-within-other-forms are symbols of intra-uterine existence or maternal protection. Alschuler and Hattwick have named the pattern "the container configuration."

The relationship to both mother and siblings was expected to emerge from the drawings based on this pattern.

4. *The single straight line pattern.* (Fig. 4). "Maleness" is considered to be another basic human experience. The reasoning in selecting "maleness" as one of the commonplace areas of experience is the same as that used in the selection of "femaleness."

Freud (3, p. 368), generalizing on the shape of the male sex organ, is responsible for the generalization that, in dreams, oblong objects symbolize the quality of "maleness." Buck (4) in a cautious statement, has also intimated that, in his H-T-P test, a tree might carry phallic significance. Alschuler and Hattwick (1), in their study, have more fully confirmed Freud's generalization by showing that straight lines or elongated figures are connected with aggressive moods of the more "masculine children." When frustrated, Alschuler and Hattwick found, children cross out the straight lines, or draw them horizontally instead of vertically. In Franck's (7) study, "masculine design" consisted of elongated forms. She found that, on the basis of round forms or straight lines, she could differentiate those who preferred the "feminine" from those who preferred the "masculine." Finally, Immergluck and Cohen (11) have found a definite relationship between stimulus words such as "father" and vertical lines.

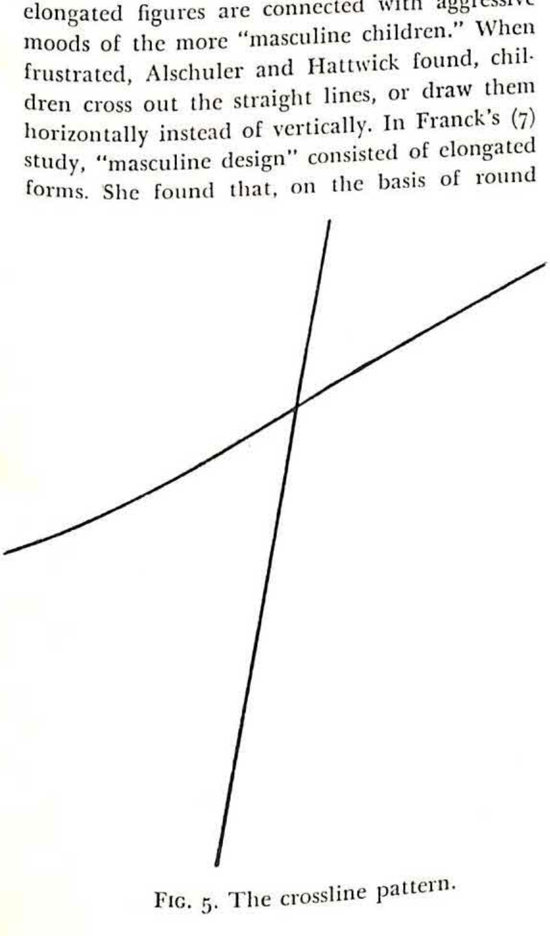


FIG. 5. The crossline pattern.

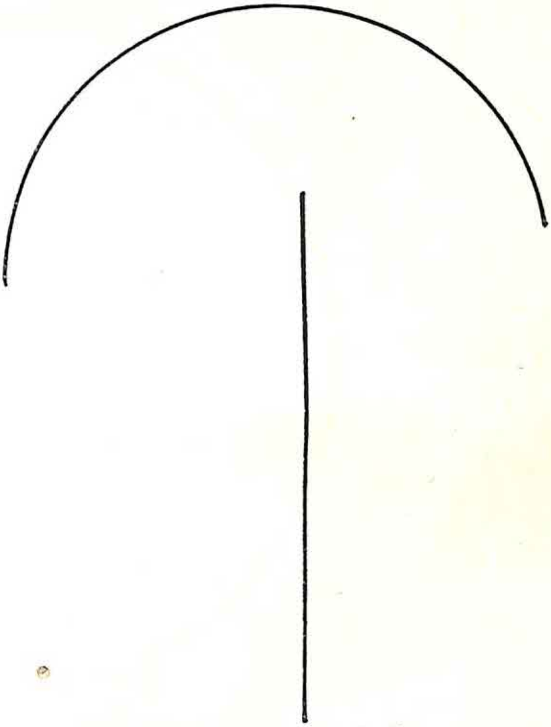


FIG. 6. The semi-circle and single line pattern.

The straight line was thus selected as the symbol for "maleness." The drawing responses based on this pattern were expected to reflect the subject's atitudes toward males and the nature of his or her experience with males.

5. *The crossline pattern.* (Fig. 5). "Male society" must be considered as basic a human experience as is "female society." The same logic as was used in the selection of "female society" as an area of common experience was hence applied to the symbolization of "male society" also.

It was considered that, if one line represents "maleness," two crossing lines might be assumed to represent some form of interaction between males. The pattern here evades the possibility of association with such consciously perceived cultural patterns as a cross or the letter "X."

If the basic hypothesis is correct, this pattern was expected to stimulate drawings having reference, in a male subject, to another man and, in female subjects, to men in general.

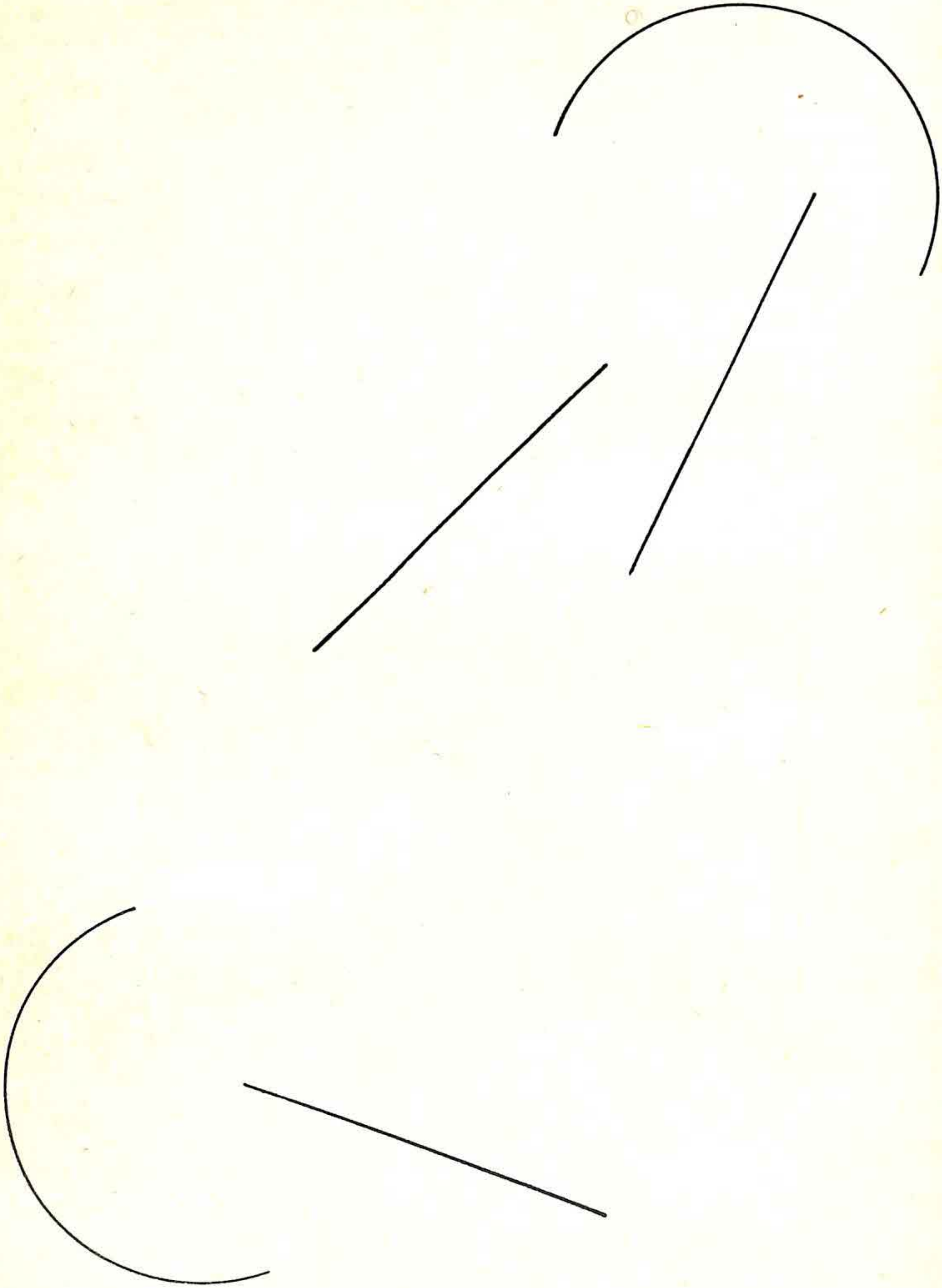


FIG. 7. The lines and semi-circles pattern.

6. *The semi-circle and single line pattern.* (Fig. 6). Interaction between the two sexes is as universal as the existence of both sexes in any society.

In order to symbolize such interaction, a combination could be made of the previously discussed symbols for "maleness" and "femaleness." The combination as given in this stimulus-pattern should imply relationships with members of the opposite sex, if the meanings ascribed to these figures *individually* are as claimed.

7. *The lines and semi-circle pattern.* (Fig. 7). The attitudes expressed in any intimate heterosexual situation are not always the same as those shown in a competitive heterosexual situation. However, the latter may also be assumed to be a common area of experience.

The field, in this case, contains several hypothetical symbols for "maleness" and "femaleness." It seemed possible that this design might, in both sexes, evoke reactions to a competitive heterosexual situation.

8. *The incomplete "U" pattern.* (Fig. 8). An attitude toward the self is another universal among human beings living in groups. A stimulus-pattern which offers an opportunity for the projection of the self-concept must be a figure which is not like the hypothetical symbol for "maleness" nor like the one for "femaleness."

The incomplete "U" pattern fulfills this description because it is composed of elongated lines which, at the same time, are somewhat curved. This enables the subject to project either pattern with equal ease. The "U" shape seemed to be such a pattern. Furthermore, this figure possesses the quality of inclusiveness which might provide a subliminal cue to the subject in answer to the question: "What is con-

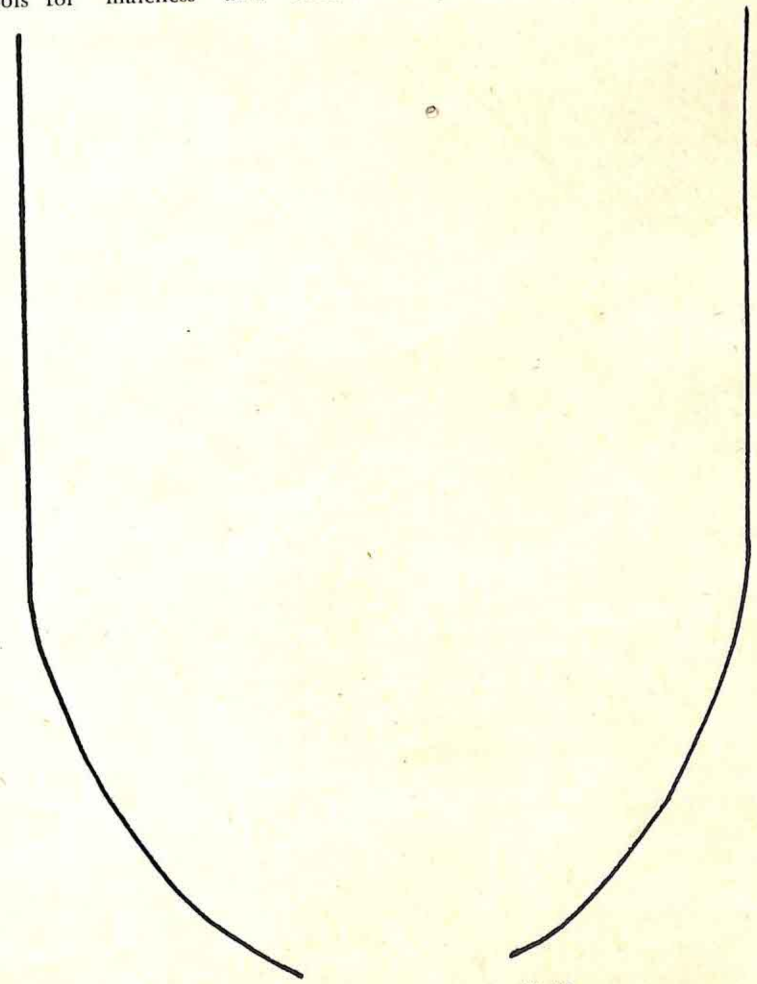


FIG. 8. The incomplete "U" pattern.

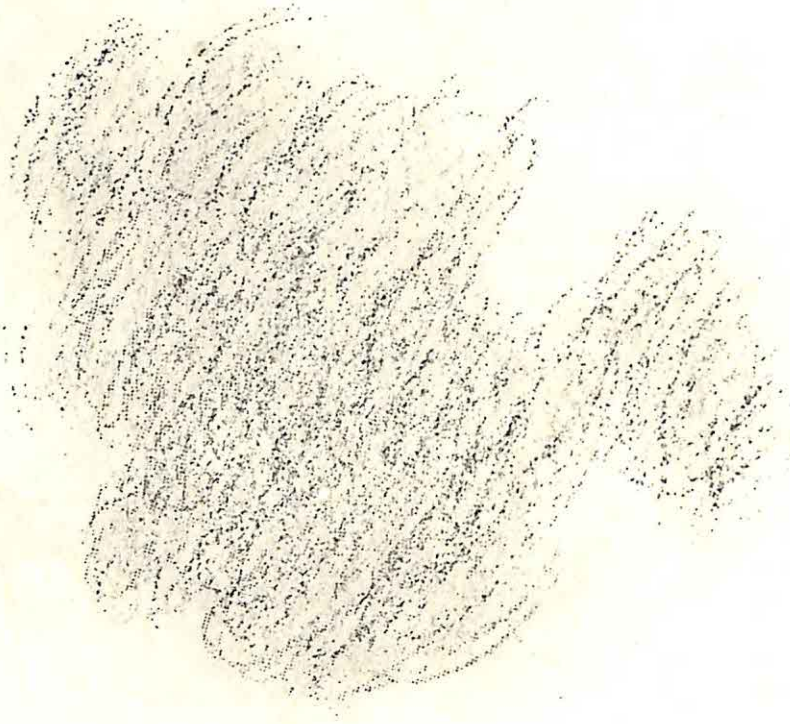


FIG. 9. The diffuse mass pattern.

tained therein?" Thus, in terms of projection theory, as applied in current techniques (see section on *Projective Techniques,* last chapter), the "U" pattern might elicit a response to the question "What is contained within you?" The subject can respond to this question apparently only in terms of himself.

The figure used here is a "U"—left open at the bottom to permit greater freedom for the subject.

9. *The diffuse mass pattern.* (Fig. 9). It is assumed that all people at times meet situations which cause uncertainty of feeling, or present a threat of some kind. This element of uncertainty is further assumed to carry with it negative affect. The writer considers this to be what is meant by anxiety. In this sense, anxiety-provoking situations are universal.

Rorschach discovered a relationship between diffuse, grey, formless stimuli on the one hand and feelings of anxiety on the other. The diffuse mass pattern was hence used here as a stimulus representing an anxiety-provoking situation.

The assumption was not that everyone suffers from a "state of general anxiety." It is the way in which the subject copes with anxiety-arousing stimuli that was expected to be ascertained by the response to pattern 9.

10. *The inverted "V" pattern.* (Fig. 10). Be-

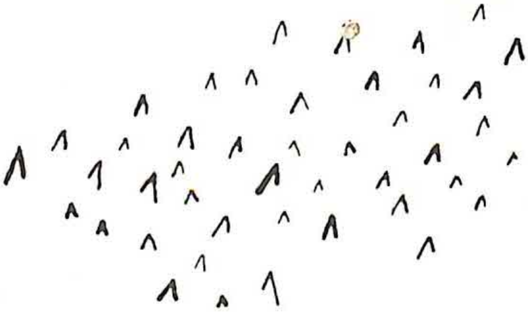


FIG. 10. The inverted "V" pattern.

cause of the importance of sex-role patterning in any society, it was assumed that there are a sufficient number of situations with sexual significance (causing uncertainty of reaction and/or evoking anxiety) for sexual anxiety to be considered a special area of human experience.

The pattern was suggested by the analysis of dreams of psychiatric patients. In their dreams, the inverted "V" pattern was related to sexual anxiety. This pattern was therefore included in the experiment to test the commonness of this association. Perhaps it is the formlessness of the pattern that is responsible for its anxiety-provoking function. Its discrete elements, which may be regarded as either masculine (angular points) or feminine (hollow figures), would probably be responsible for any sexual connotation it might have.

As in the case of general anxiety, the interpretation here was in terms of the individual's manner of coping with such situations as tend to provoke specifically sexual anxiety.

11. *The jagged-line pattern.* (Fig. 11). Hostility and aggression may be assumed to be a common experience of human individuals in our society, and in a multitude of societies other than our own.

The study by Immergluck and Cohen (11) confirms the plausibility of the assumption that the jagged-line figure evokes responses of aggression. These investigators have reported that jagged lines were frequently associated by their subjects with such words as "war" and "aggression." Dream material from psychiatric patients also confirms this assumption. The association between a jagged line and the concepts of speed or force might be responsible for the tendency of this pattern to disclose hostile or aggressive impulses.

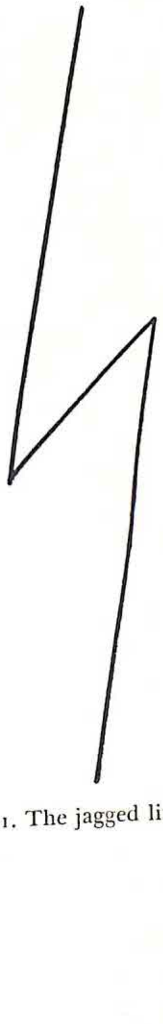


FIG. 11. The jagged line pattern.

Chapter II

PROCEDURE

A. Presentation of Stimuli

The first step in approaching the subject was to make it clear to him that what he was about to do was in no sense a test of his art ability. When pressed for an explanation, we have occasionally told subjects that the procedure was a "kind of test of ingenuity." However, almost always subjects were content with the instructions as given. The usual statement was as follows:

> You will notice that the first sheet here has nothing on it.[1] You may draw anything at all with these pencils. There will be no time limit; take as long as you wish. When you have finished the first drawing, turn the first sheet and take the second sheet by this corner (demonstrating) and flip it over. [This procedure was to accomplish two things: one, to prevent the subject from seeing the stimulus-pattern before he was ready to draw upon it; and, two, to make certain that the pattern was seen in the same position by all subjects.] On the rest of the sheets, you will notice there are some lines already present. You may again draw anything that suits your fancy, but on these sheets try to include the lines, if possible.

The order of presentation of the stimulus material in two-thirds of the cases avoided placing a stimulus-pattern immediately after one with a related hypothetical meaning. For example, the "maleness" pattern (Fig. 4) was separated from the "male group" (Fig. 5) pattern. This was to prevent carry-over of ideas which might make it difficult to establish the stimulus-pattern to which the subject was reacting. The one purposeful exception was the placing of the sexual anxiety stimulus before the one for general anxiety. It was feared that otherwise, especially in our society, a subject might react to the general anxiety stimulus in sexual terms, when he could react in other terms if first given a chance to "drain off" sexual anxiety. There remains a question of whether placing the sexual anxiety stimulus first might not arouse feelings which would then artificially influence performance on the general anxiety pattern. The writer believes that the answer lies in the release of tension provided by an expression of attitude or "catharsis." Since the subject was not merely expected to respond to certain stimulus-patterns (as in Rorschach or TAT) but given an opportunity for individualized expression by even changing the stimuli, it was felt that a certain order of presentation of the stimulus-patterns was indicated. To justify this order, and as a control, random order was used in the last third of the cases.

When the drawings were completed, a sheet of paper bearing the inquiry questions was presented with enough paper for the subjects to answer each question fully for each symbol. The questions read as follows:

> Question A. What is this a picture of?
>
> Question B. Was it relatively easy or hard to draw?
>
> Question C. Do you like it?
>
> Question D. What mood does it represent?
>
> Question E. (1) Describe in *detail* the personalities of any people (or animals) depicted. (2) What are they doing? (3) What are they thinking and feeling?
>
> Question F. What mood do you feel looking at it?
>
> Question G. What does it remind you of?

---

[1] In the presentation of the stimulus-patterns, each subject was given a blank sheet of paper on which to make a "free drawing." The purpose of this was to offer an opportunity for expression of attitudes which were brought into the experimental situation. Analysis of these drawings is not necessary to test the hypothesis of universal symbolism.

The average amount of time taken for the entire procedure by each subject was one hour.

When it was possible to do so, notes on the manner in which the subject worked were appended to each record. For example, when the subject became suddenly tired while working on one symbol and expressed the fact to the test supervisor, the matter was noted, particularly in terms of the symbol upon which the subject was working at the time. For most of the cases used in this study, of course, it was not possible to make such observations because of the difficulties encountered in group administration.

When the drawings and answers to the inquiry questions had been collected, the next step was to analyze them in terms of the hypothetical meanings of the stimulus-patterns. The section following describes in detail what this process involved.

B. Analysis of Stimulus Material

Three major sources of data were available for the interpretation of these stimulus-patterns. The first was the *manner of drawing;* the second, the *content of the drawing;* and the third, the *verbal associations* as revealed by the subject in statements on the inquiry.

The *manner of drawing* on the stimulus sheets was not given systematic emphasis in the analysis of any single drawing by itself. Valuable information about a total personality may be gained by the application of free-drawing-analysis techniques to the manner of drawing. In this study, however, a complete understanding of the personality of each subject was not required. For this reason, the manner of drawing was analyzed only when there seemed to be outstanding differences between one drawing and the rest of the subject's set of responses. When this occurred, such matters as the use of color, the "goodness" of a particular drawing in relation to the quality of drawing on the other sheets, and the general *Gestalt* of the drawing, i.e., whether there had been closure of the figure and whether the whole sheet was used, etc., were taken into account. In this context, such factors as these helped clarify the subject's responses to the individual stimulus-patterns.

The *content of a drawing* is often obviously related to the hypothetical symbolic meaning of the stimulus-pattern. Thus, for example, one subject drew a baby-in-bunting out of the concentric semicircle pattern, and labelled it "A baby which unfortunately looks like an adult." The implication here is that the subject acknowledges mother-dependence and yet accepts status as an adult. The inquiry, of course, was of the essence. But even where no inquiry was conducted (as in the group of normal children), the subjects often spontaneously labelled their drawings. When they did not do so, the content was sometimes so obscure that it was impossible to tell what it was, or what it might mean to the subject. At times, of course, it may be possible to make some interpretation about the personality from the content of the drawing alone, even when the content is not quite as obvious as in the example cited at the beginning of this paragraph. For instance, when a drawing is labelled, but no further inquiry is available, the following might still be interpretable at the risk of subjectivity. For example, lollypops might be presumed to imply positive affect; targets might be said to imply goal-direction; moons might be presumed to connote remoteness; etc. Such interpretations, however, offered without the subject's verbalizations, could never be made without misgivings.

Answers to the *inquiry questions* provided *verbal associative material.* Consequently, when answers were available, as they were for most cases in the study, they made interpretation defensible. The answers to the questions as to what the drawing was, whether it was easy or difficult to draw, and whether or not the subject liked it, details about the "personalities" depicted, and any other material which the subject might associate with his drawing, were straightforwardly given by most subjects. The replies to the questions on liking the picture and ease of drawing it were always usable for further interpretation. Various other replies could be analyzed in terms of the subject's liking or disliking of the picture and his ability to draw easily when presented with a particular pattern. The possibility suggested itself here that a subject might like a particular stimulus-pattern, but still not draw well upon it, nor like his own product. When this happened, subjects were likely to specify that they disliked their execution of the pictures, but not the concepts they had in mind. The subjects themselves thus helped in the interpretation of the inquiry.

The two questions about "mood" (mood of the picture and mood it evoked in the subject) provided further information about how the subjects evaluated their own responses to the stimulus-patterns. It is interesting to note that the responses to the two questions about mood were usually not the same, although they were most often complementary. An example of this is the statement that the mood of a given picture was "horror," while the feeling evoked in the subject was one of "considerable discomfort." When the subject is in conflict about the symbolized area, the various inquiry questions usually stimulate different affectual tones in the various replies. For example, the mood of a given picture could be "horror," while the feeling evoked in the subject might be described as one of "fascination." The answers to the questions about liking the picture and the ease with which it was drawn are basic to the rest of the analysis.

The method of analysis must also be understood in terms of the differences of approach required by the different stimulus-patterns. Following is a discussion of the analysis in terms of each stimulus-pattern.

The *semi-circle pattern* (Fig. 1) has been assumed by other investigators to evoke responses to "femaleness." In order to make specific interpretation of a subject's responses to the symbol of "femaleness," it is, however, necessary to establish a relationship between the concept and certain areas of the person's developmental history. The assumption was that responses to the concept of "femaleness" could be proved to be based upon contact with members of the female sex. It was further assumed that the subject's concept of "femaleness" could be shown to be influenced by his or her earliest or strongest experiences with a member of the female sex. Since, for most people, this member would be the mother and/or mother-surrogate, the interpretations based upon the drawings on the semi-circle pattern were taken to relate to the mother and/or mother-surrogate.

Similar logic applies to the straight-line pattern (Fig. 4). This pattern is related to "maleness." Here, as in the case of the symbol for "femaleness" (Fig. 1), it was assumed that the subject's modification of the symbol might be in terms of his generalizations from the male figures who had made the strongest impression upon him. In most cases, this would be the father and/or father-surrogate.

In connection with the figures for "maleness" (Fig. 4) and for "femaleness" (Fig. 1), there are two points of theoretical interest. The first is the extent to which identification on the part of the subject with either parental figure might influence his drawing. The self-concept might thus enter to a degree. In other words, the subject might identify with the one or the other of his parents. Then, in reacting to the appropriate stimulus-pattern, he might indicate characteristics which are true of himself, but not of the parent. When this happens, however, he still reveals his concept of "maleness" or "femaleness," as the case might be. For example, a boy who had identified with his father drew a "scale of justice" out of the straight line. On the inquiry, he stated that it was easy to draw, that he liked it, that its mood was "idealism," evoking in him a mood of "determination." Finally, it reminded him of the "frailties of man."

The interpretation was that there was a positive relationship between the subject and his father, and that the subject somehow derived from his father his ideals and ambitions. From the last association on the inquiry, however, it seemed that the father probably had shown some significant weakness which was important to the boy.

Case history material brought out the following facts: When the boy was a year old, his father was dismissed from the police force for bribery. The relationship between father and son was a good one, and the boy adopted his father's walk and manner of speech. At the time of testing, when the boy was seventeen years old, he expressed interest in law with the idea of regaining honor for his family. He was already well known in his comunity for his honesty, forthrightness, and high ideals of citizenship. Thus it may be seen that a subject may show clear identification with the parent and on the appropriate symbol include his own traits as well as information about the parent.

The second point of theoretical interest concerns those cases where the subject drew male figures on the semi-circle (Fig. 1) and female figures on the straight line (Fig. 4). The assumption made here was that the subject was still expressing his concept of "maleness" or "femaleness," but that he viewed these in terms of reversal of the usual roles. Thus the drawing of a man's face on the semi-circle has been shown in various cases to be related to experience with a domineering mother.

At first the writer considered the possibility that a person might respond to the "femaleness" pattern (Fig. 1) in terms of a male figure because of association with weak males. Similarly, it did not appear certain that a female figure drawn on the straight line referred to weak males rather than to strong females within the realm of the subject's experience.

The inquiry material resolved this dilemma. It was usually clear from the answers exactly which parental figure was involved. For example, one subject drew a grim-looking bald-headed man out of the semi-circle (Fig. 1). The straight line became part of a hotel described as being in total darkness because all of "the lights are out, and everyone is over at the tavern across the street." Further, the bald-headed man was called a "business man who knows his stuff," and while the mood evoked in the subject was one of "admiration," yet he stated that he did not like the picture. On the other hand, the mood evoked in the subject by the straight-line drawing was that of "longing," and he expressed a liking for the picture.

The case history showed that the mother was a strong, determined woman who managed the family budget. The father was characterized as a relatively weak man who worked as a traveling salesman, and tended, when not controlled by his wife, to seek satisfaction in drink. The subject was described as being more fond of his father than of his mother. However, the father's occupation prevented close companionship with the subject.

This kind of check, bolstered by numerous examples, lent strong support to the hypothesis (see case analyses and validation data) that the semi-circle (Fig. 1) is related to "femaleness."

The *crossline pattern* (Fig. 5) and the *counter-opposed semi-circle pattern* (Fig. 2), because they both represent plural situations, would seem to focus the subject's attention upon social aspects of his attitudes toward "maleness" and "femaleness," rather than his attitudes toward individual figures who might be prototypes of his concepts of "maleness" and "femaleness." Thus, the interpretations made (see validation material and case analyses following) on the basis of the drawings evoked by these two stimulus-patterns relate to the subject's attitudes toward the male and female members of his group, as he feels toward them "en masse," so to speak. The kind of social technique which an individual may use often varies with the sex with which he is dealing. This kind of difference stands out clearly in the drawing material and, when it occurs, note is taken of it in the analysis. For example, on the *counter-opposed semi-circle pattern* (Fig. 2), one subject drew a line connecting the two arcs at the top. This he called a table. The *crossline pattern* (Fig. 5) he turned into a pair of sabres held by two opponents. From the content of the drawing and the inquiry material, it was concluded that the subject considered himself to be a source of support to women in the "strong silent" manner. The manner of drawing was of help in making this interpretation also, because it has been observed that use of the space between the areas seems to be characteristic of those who wish to establish that they are not identified with the female group. The interpretation on the second drawing was that the subject was generally antagonistic to those of his own sex. Case history data, and the information given by the subject's psychiatrist for validation purposes, confirmed these interpretations.

There are two stimulus-patterns that are combinations of the male and female symbols. The first one to be considered here consists of *one semi-circle and one straight line* (Fig. 6). Being a combination of the male and female symbols, this pattern, in the relationship represented, appeared as if it might have relevancy to the intimate heterosexual situation (see case analyses and validation data). The kind of specification on this symbol that a given subject makes depends presumably upon the range of his experience and often on his age. The subject may or may not relate this stimulus-pattern to experiences of physical contact. The subject's interpretation tends to be in terms of his attitude toward a close relationship between two people, one of each sex.

The second combination consists of an arrangement of more than one of each sex-symbol in the field (Fig. 7). This arrangement is such as to suggest competition between members of the same sex for those of the opposite sex (see tabular material on this pattern). Again, the degree to which sexuality, in the sense of physical contact, is involved depends on the subject.

An example of how the attitude toward the intimate heterosexual situation may vary from the one expressed for the competitive heterosexual situation is the following: A subject drew on the intimate heterosexual stimulus-pattern what he called "a balloon drifting lazily toward the earth." On the competitive heterosexual stimulus-pattern (Fig. 7) he drew a "dart game in which each guy sees who can get more." The interpretation was that the subject felt stimulated by competition with other men for women, but his aggressiveness did not extend to the intimate level, where he developed a more passive attitude. This analysis was in accord with the subject's own appraisal of his behavior. Of course, other subjects show the opposite attitude, withdrawing from the competitive heterosexual situation, but becoming assertive when secure in an intimate heterosexual situation. Most often, the attitudes expressed on the two stimulus-patterns are complementary.

The *incomplete "U" pattern* (Fig. 8) permits of ready projection of a face, but also of a different use, if the subject is not inclined to draw a face. This seems to offer an opportunity for the expression of attitudes relating peculiarly to the self. Again, the method of modifying the symbol is taken into account in the analysis. However, particular stress is laid upon the content of the drawing in making interpretations as to the self-concept.

The most frequent response is the drawing of a vessel of some sort, usually a glass or vase. Such containers are passive receptacles, and imply an attitude of passivity on the part of the subject. At least, we have so interpreted them in our analyses (check against tabular material on the pattern involved). A striking example was the drawing of one male subject who labelled the main figure a pot, drew a fire underneath, and stated that it was the "common pot for all the bums who keep taking from it." The case history shows that the subject was a member of a large family in which he was known as the "good one." He was considered to be the person to ask for any favors. Particularly interesting was the fact that each member of the family said that he was shamefully imposed upon by all the others, but no one admitted imposing upon the subject. Living with his younger brother in college, he did the housework and cooking for both; and also lent his brother much money from his own government allowance, even though his brother was receiving the same monthly amount. At the time the drawings were administered, the subject was seeking psychiatric aid in adjustment to marriage. He felt that advantage had been taken of him in arranging the marriage, and he felt that he "suddenly couldn't take it any more." These facts would seem to confirm the interpretation of his drawing, which was that the subject felt himself to have played a passive role in which others had taken advantage of him. Not all subjects who draw passive figures on this stimulus-pattern indicate conflict about it, of course. Also, drawings other than those of vessels may indicate a passive element.

Often faces are drawn, and in the inquiry they may bring forth any emotion in the human range. The guiding principle in interpreting reactions to the incomplete-U pattern is that the self-concept alone is revealed. If an individual has misconceptions about himself, they will be in evidence. Thus, a strikingly beautiful girl with an I.Q. of 152 (Wechsler-Bellevue) and various college honors, including Phi Beta Kappa, said on the inquiry about the face she had drawn on the incomplete-U pattern (Fig. 8): "She is silly, shallow, vain, and quite unworthy—not even really pretty." When asked for her real opinion of herself during the interview this subject spent eight minutes telling why she was not actually as personable as most people seemed to believe. Naturally, misconceptions about the self can be in the direction of self-flattery, also. In any case, it is how the subject regards himself that is sought in the validating data for the incomplete-U pattern interpretations.

In the symbols for sexual anxiety (Fig. 10), general anxiety (Fig. 9), and aggression (Fig. 11), the analysis was focused on two possibilities: First, and most important, the manner of handling, i.e., the degree of structuralization added by the subject, and the way in which this was accomplished. By way of example, we might take the subjects who do not draw anything on these stimulus-patterns. Failure to use this pattern might mean that, when confronted by an anxiety-provoking situation, the subjects could not focus on them even though told to try to do so. The analysis, then, should be that the subject could not deal effectively with the type of situation symbolized.

In other cases, the subjects made a stimulus-pattern an integral part of a more complete drawing. Thus, the aggression-stimulus (Fig. 11), might be used as the side of a building. In such an instance, the analysis would be that the subject reacts to aggression situations constructively, as much because of the constructive use of the stimulus-pattern as through the content of the drawing.

The second aspect of the analysis referred to the content of the drawings made on these three stimulus-patterns. It was felt that the content might reveal the kind of experience the subject tended to relate to the stimulus-pattern. Thus, a subject who called the general anxiety stimulus-pattern "dirty combings" was considered to be concerned with matters of cleanliness. On a personal inventory form, this subject indeed stated: "Cleanliness is next to godliness." She was under treatment for compulsive symptoms, including compulsive washing.

The semi-circle within a larger semi-circle pattern (Fig. 3), or regression symbol, relates to the subject's attitudes toward the mother. However, it is especially important in revealing the subject's security strivings as contrasted to a sense of emotional independence. All interpretations contain reference to this point. The *manner* in which the subject handles the stimulus-pattern is frequently important in analyzing responses to this stimulus-pattern. For example, it is possible to "close" the figure in different ways. One way is to complete both circles (e.g., automobile tires). Regardless of the implications of the content, completion of the circles in this manner could be interpreted to mean a dependency (regressive) reaction. The logic here is that the structuralization involved tends to emphasize the circle-within-a-circle pattern. Some subjects connect the parallel ends of the two circles (e.g., a "bitten-off doughnut"), emphasizing emptiness. Here interpretation must take into account the fact that the subject has given up expecting maternal protection. Emotionally independent, he feels free to manifest aggressive behavior. This kind of analysis has been substantiated by the validating material.

The importance of the content is evident in this pattern also. The inquiry material often sheds light on sibling relationships as well as the attitude toward mother dependency. One subject, aged eight, made a "flat car tire." On the inquiry he made it clear that this had been caused by a sudden deflation. The interpretation was that he had been relatively secure and dependent in attitude, but that a recent shift in interpersonal relationships (probably with his mother) had suddenly left him feeling insecure. Case history material showed that his first younger sibling had been born three months prior to the time of drawing. As the youngest member of the family until this happened, the subject had been the "family pet," and used to spend most of his free time trying to help his mother with her housework. He was described as bewildered by the sudden loss of constant maternal attention. These facts support the interpretation of his drawing.

Further examples of the method of analysis are given in two complete cases included in a later chapter.

C. VALIDATION DATA FOR EACH SUBJECT GROUP

This study comprises data on 169 subjects. Of these, 157 are white Americans and 12 are American Indians. They are almost evenly divided as to sex, and range in age from six to sixty-nine.

In all cases, blind analyses of the drawings were made first. Then, specific interpretations were checked against validating data. The sources of validation varied with the group, and are completely described in this section. The same rating system was used in all groups, however, to express degree of agreement between the analyses and the validating data. This system is fully described in the following sections. It provided a checking system in terms of the following possibilities: *Complete Corroboration, Partial Corroboration, Conflicting Evidence, Partial Conflicting Evidence, Complete Contradiction, Partial Contradiction,* and *No Evidence.*

The subject groups and data available for checking the validity of drawing interpretations follow:

Group No. 1: NORMAL CHILDREN

Description: This was a group of 30 public school children, evenly divided as to sex, and all approximately 13 years old, considered a part of a "normal" population.

Method of Validation: This group has been the object of concentrated study as part of a program of cooperative effort by psychologists, social anthropologists, sociologists, and other members of the Committee on Human Development at the University of Chicago. The various workers, operating independently, employed different techniques of study. The kinds of materials which were utilized and analyzed independently by the various members of the conference team were as follows: case history material; interview data obtained from parents, teachers, peer groups, and subjects; socio-economic class ratings; sociometric data; Rorschach test; Thematic Apperception Test; personal inventory tests, such as the California Personality Test; essays on such topics as "The Person I'd Most Like to be Like"; and material on the development of moral conscience. A year after investigations were completed, the stimulus-material for this study was administered. A clinical case conference was held in each case, at which time there was a critical examination of the findings obtained by each technique (excluding the present ones). The summary transcripts of these conferences were used as validating data on the present group of cases.

The drawings completed by the children were analyzed without knowledge of more than their age and sex. Inquiry questions were not administered to this group. The interpretations derived from the drawings were rated with reference to degree of agreement with the clinical case conference conclusions. In each case, the investigator compared her interpretations with the independent clinical summaries. Each point of the present diagnoses was rated on degree of agreement with the validating data. The rating system is described in a later section.

Group No. 2: NORMAL ADULTS
SUBGROUP A—VOCATIONAL GUIDANCE GROUP

Description: There were two subgroups making up the normal adult population of this study. One consisted of people who applied for vocational guidance at a general psychological clinic. There were 13 men and 11 women, ranging in age from 17 to 61 years. None of this group of subjects exhibited a psychiatric syndrome, either clinically or in test findings. They are consequently regarded as members of an essentially "normal" population.

Method of Validation: All members of this subgroup were given a battery of tests investigating personality structure, including: the Rorschach Test, TAT, Szondi Test, and the California Personality Test. In addition, there was a prolonged conference between the subject and a professional psychologist, during which the test results were discussed. This psychologist used the test findings and his notes on the conference as the basis for his rating of the writer's blind interpretations. As in the other groups, each point of interpretation was compared with the validating data and rated on degree of agreement with these data.

Group No. 2: NORMAL ADULTS
SUBGROUP B—VOLUNTEER NORMAL ADULT GROUP

Description: The second subgroup consisted of 30 adults, evenly divided as to sex, and ranging in age from 18 to 49 years. They had all had some amount of college training, and all had volunteered to take part in this study. None of this group had ever been required to seek psychiatric aid, and therefore they were considered to be members of a relatively "normal" population. Some members of this group were doing advanced work in psychology. None of them, however, indicated any familiarity with the theoretical implications of the drawing-task.

Method of Validation: In many cases of the other groups (such as 3 and 4), the interpretations were concerned with material which, because of their emotional conflicts, the subjects could not be expected to discuss. It was, therefore, considered desirable to have a group of normal subjects for whom conscious acceptance of the interpretations might be the criterion for validation. The analyses were written in a manner providing specific points for judgment. Generalities such as "manner of handling anxiety" were carefully avoided. As much as possible, the subjects were asked non-leading questions, and encouraged to elaborate upon their answers.

Group No. 3: CHILDREN WITH BEHAVIOR PROBLEMS

Description: This group consisted of 13 children who were under treatment for behavior disorders. They ranged in age from 6 to 16 years, and were nearly evenly divided as to sex. Their symptoms ranged from relatively mild to relatively more severe types. As in the case of the adult psychiatric group (see below) no special account was taken of the individual diagnoses. These subjects were being treated by two therapists.

Method of Validation: The list of interpretations based on the drawn material was submitted to the therapist treating the case. The therapist, using his clinical notes for reference, scored the interpretations in terms of the system previously described.

Group No. 4: ADULTS IN PSYCHOTHERAPY

Description: This group consisted of 60 adults, men and women, ranging in age from 17 to 69. All of them were under treatment for psychiatric disorders at the time the drawing material was administered. These subjects were treated by three therapists. The diagnoses and symptoms of these patients varied from the mild neuroses to diagnosed schizophrenia. However, for the purpose of this study, no differentiation was made among the members of this group in terms of types or symptoms. The relationship between specific disease types and responses to the stimulus pattern must be left to future research. The purpose of including these subjects is to test the hypothesis that the symbols are meaningful to people with emotional aberrations in the same way as they are to others.

Method of Validation: The list of interpretations based on the drawn material was submitted to the therapist treating the case. The therapist, using his clinical notes for reference, scored the interpretations in terms of the system previously described.

Group No. 5: INDIAN GROUP

Description: This was a group of twelve children, all approximately thirteen years old. There were eight girls and four boys. Seven of the children belonged to the Navaho, Zuni, and Papago tribes. These are all matricentric tribes. The others belonged to the Sioux tribe, which is patricentric. There are other differences, in that the Sioux are considered to possess a disintegrating culture, while the others are not.

Method of Validation: These children were part of a group studied in 1943 in the Research on Indian Education of the Committee on Human Development. At that time, material was gathered in the form of autobiographical notes, case histories, TAT, Rorschach tests, and free drawings. The process of using this material for validation purposes was the same as in the case of the normal children's group.

D. SYSTEM FOR VALIDATION-RATING

Seven rating categories were used to indicate the relationship between the interpretations on the drawing materials and the validating material. As previously noted, these were *Complete Corroboration, Partial Corroboration, Conflicting Evidence, Partial Conflicting Evidence, Complete Contradiction, Partial Contradiction,* and *No Evidence.*

The same system was used for all the groups regardless of the source of validating material. Fuller explanations of the rating categories are given below. Examples taken at random from several different cases are included. These cases are all from the Normal Children's group. It will be recalled that validating statements were available for direct quotations from the transcripts of Rorschach, TAT, and other test records, and from conclusions of independent workers in a clinical case conference on each case.

Plus (+) was used to indicate *complete corroboration* of the interpretation by the validating material. Only when the nature of the evidence consisted of an identical paraphrase of statements used in the blind interpretation was this scoring used.

Examples of Plus Scoring

1. Interpretation:
   "Handles anxiety generally through evasive generalization."
   Validation:
   "Anxiety motivates over-all adjustment pattern. Seen in all projective materials in a hesitancy, evasion, and cover-up pattern." (Summary transcript of clinical case conference)
2. Interpretation:
   "Sexual anxiety controlled: fantasy methods."
   Validation:
   "Sexual anxiety controlled through fantasy." (Summary transcript of clinical case conference)
3. Interpretation:
   "Finds security in companionship with those of same sex."
   Validation:
   "Personal achievement (sports) closely tied to sense of loyalty to group and need of recognition from the group."
   "Close identification with young men."
   "Plays with older boys on equal terms."
   ". . . He is aware of masculine peer group expectations and trying to live up to them." (All the above quotations are from the clinical case conference)
4. Interpretation:
   "Self-concept: very masculine."
   Validation:
   "Very high masculine score." (Masculinity-femininity profile)
   "Strong male self-concept." (Essay material)

Partial-Plus (P+) was used to indicate *partial corroboration,* that is, partial evidence of a corroboratory nature. In all cases where there was some question as to whether a "+" or a "*P+*" was applicable, the "*P+*" was used.

Examples of Partial-Plus Scoring

1. Interpretation:
   "Playful attitude as method of adjusting to competitive heterosexual relations, always with controlled attitude."
   Validation:
   "Satisfactory adjustment to heterosexual

role on interest level." (Interpretation of California Personality Test in clinical case conference)

2. Interpretation:
   "Feels mother-rejected."
   Validation:
   "Mother impersonal." (Summary and interpretation on interview material in clinical case conference)
3. Interpretation:
   "Oral fixation."
   Validation:
   "Breast-fed for almost a year." (Case history)
4. Interpretation:
   "Heterosexual relations on a close basis are repulsive and threatening."
   Validation:
   "Possible passive homosexual adjustment." (TAT)
   "Low heterosexual interest." (TAT)

Plus-Minus (±) was used to indicate *conflicting evidence.* Sometimes, in cases where the validating material consisted of compilations of reports on the subjects by different workers, the amount of information present in the record was not sufficient to explain seemingly conflicting results from the different techniques employed. In such cases it was assumed that we could not judge whether the evidence supporting or contradicting our interpretation was correct. The "±" sign was used to denote this condition.

Examples of Plus-Minus Scoring

1. Interpretation:
   "Positive attitude toward individual heterosexual contact."
   Validation:
   "Seeks transference from mother to 45-year-old neighbor woman with whom he goes fishing." (Case history interpretations in clinical case conference)
   "Not interested in girls." (Case history)
2. Interpretation:
   "Hostility freely expressed."
   Validation:
   "Characterized by temper outbursts." (Case history)
   "Considered submissive and cooperative." (Sociometric)

Partial-Plus-Minus (P±) was used to indicate *partial conflicting evidence,* that is, partial evidence which at once supported and contradicted the interpretations, when this evidence was inconclusive in both directions.

Examples of Partial-Plus-Minus Scoring

1. Interpretation:
   "Hostility toward mother."
   Validation:
   "Mother torn between need for control and laxness because of subject's charm." (Case history interpretation in clinical case conference)

Minus (—) was used to indicate *contradictory evidence.*

Examples of Minus Scoring

1. Interpretation:
   "Expects coldness from girls."
   Validation:
   "Girls show interest in him." (Sociometric)
2. Interpretation:
   "Mildly compulsive in control of sexual anxiety: chief source of control in attitude of repulsion."
   Validation:
   "No sexual anxiety." (TAT)

Partial-Minus (P—) was used to indicate *partial contradiction,* that is, evidence of a contradictory nature.

Examples of Partial-Minus Scoring

1. Interpretation:
   "Experimental mood with regard to competitive heterosexual situation."
   Validation:
   "Prognosis toward satisfactory heterosexual adjustment not favorable." (TAT)
2. Interpretation:
   "Sees male figures as sources of support."
   Validation:
   "Does not play with boys of peer group very much." (Case history) (No information available on father or other non-peer males)

The category, *No Evidence (O),* was used to indicate lack of evidence in either the positive or negative direction as to the validation of the interpretation. It was used when no statement could be found anywhere in the case record which shed light even remotely upon the hypothesis scored. As the examples above should make clear, conservatism was the rule throughout, and any statements which seemed to modify the scoring at all were included. This symbol was also used when the subject did not answer the inquiry questions, and the content of his drawing was obscure.

Further examples of how this scoring system was used may be found in Case A in the Appendix.

E. CHECKS FOR RELIABILITY

The checks for reliability of procedure varied with the subject groups because of the differences in method of validation. For this reason, the following discussion has been organized in terms of the subject groups. Subsequently, reliability checks on the methods of this analysis will be discussed.

The Normal Children Group.—In the validation process for this group the writer selected statements from the validating data which seemed to her to have a bearing on the specific interpretations made of the subjects' responses to the stimulus material. This raised two problems of reliability. The first is whether the writer's selection of validating material was reliable; and the second is whether the rating made on the extent of agreement between the interpretation and the validating statements was reliable.

The reliability check of the writer's selection of validating statements from the validation data was made as follows: Questions were formed out of the interpretations of fifteen of the cases. Care was taken that the questions not be "leading." An interpretation such as "Hostility freely expressed" was changed to "How is hostility handled?"; "Mother affectionate" was changed to "How does subject interpret mother's attitude?" An independent worker was then given these questions as set up, and asked to find, from the complete clinical case conference data which had been gathered, statements, if any, answering the questions.

As seen in Table 1, the checker's material was identical with the writer's in 84% of the cases. The reliability of the validating statements is underscored by the remaining breakdown. The cases in which the checker found evidence, but the writer had listed none, was 2%.

The reliability check on the writer's ratings was to give another independent worker the interpretations made on each subject, with the validating statements selected from the clinical material by the writer, and to ask her to make her rating independently. Her scores were then compared with those made by the writer. Complete agreement was found in 94% of the 76 scorings which had been calculated. Differences based on more than a half scoring value[2] were found in only 0.7% of the comparisons.

TABLE 1

RELIABILITY OF VALIDATING STATEMENTS
(Based on 215 Judgments)

Category	No.	%
Checker cited same material	181	84
Checker found different material	4	2
Checker found evidence where writer listed none	4	2
Checker found no evidence where writer did	26	12
Total	215	100

Adult and Problem Children Groups.—In the Normal Adult-Vocational Guidance Group, the Psychiatric Case Adult Group, and the Problem Children Group, validation ratings of the interpretations were made entirely by independent workers. The reliability check made for

---

[2] "A half scoring value" refers to the relationship between partial evidence and complete evidence, e.g., "*P*" and "+."

TABLE 2

RELIABILITY CHECK BY THERAPISTS

Group	% Agreement	% Change, ½ Step	% Change, Full Step	% Change to or from "0"
Psychiatric Adult Group:				
Therapist A	98	1	0	1
Therapist B	96	2	1	1
Therapist C	82	5	4	9
Normal Adults, Vocational Guidance Group	93	7	0	0
Problem Children's Group:				
Therapist A	98	1	0	1
Therapist B	97	2	0	1

these groups, therefore, was to have the same workers re-rate a representative sampling of each group after two months. Table 2 summarizes the results. The percentages on this table express the ratio of agreement between the first and second ratings made in the several groups. These ratings were made by three therapists for the Psychiatric Adult Group, a clinical psychologist for the Normal Adult-Vocational Guidance Group, and two therapists for the Problem Children Group.

For the Psychiatric Adult Group, the percentages of agreements between the first and second scorings were 98 per cent, 96 per cent and 82 per cent respectively. The last figure is lower than the others because this therapist made more "no evidence" scoring the second time. He also completely reversed his earlier scoring on some interpretations. Even so, these high percentages of agreement permit confidence in the reliability of the validation processes used for this group.

For the Normal Adult-Vocational Guidance Group, the first and second scorings were identical for 93 per cent of the interpretations. In the remaining 7 per cent, there were no shifts greater than a half scoring value. These ratings are therefore considered to be thoroughly validated for the Normal Adult-Vocational Guidance Group.

For the Problem Children Group, agreement between the first and second scorings was 98 per cent and 97 per cent respectively. This extraordinary reliability is probably explained by the fact that, in children, such factors as "oral-fixation" or "peer-group relationship" are more readily checked.

The Normal Adult Group was divided into two subgroups. One was the vocational Guidance Subgroup, validated by a clinical psychologist on the basis of clinical interviews, and standard test results. The second group was interviewed by the writer, and thus validated directly. Reliability of ratings for the latter subgroup was checked in the following manner: The percentages were computed for the number of interpretations of subjects' responses rated in the various categories. These figures were then compared for the two subgroups. Critical ratios were calculated for the percentages[3] in the "complete corroboration" category. The ratios ranged from 0.23 to 0.47. Since 2.00 is the usual criterion of significant difference, these results indicate that there is very little difference between the two validation results. For this reason, the two subgroups are considered together.

---

[3] The percentages were expressed as proportions for this purpose.

Indian Group.—Since the validation method for the Indian Group was the same as that for the Normal Children Group, no new reliability checks seemed necessary.

Method of analysis.—Two checks were made on the reliability of the method of analysis. The purpose of the first check was to discover the ease with which another worker might learn to analyze this kind of material. Ten records were given to this worker, to whom the frame of reference used in the study was carefully explained. He was asked to analyze these records, and his analyses were compared with those made independently by the writer. The essential points were analyzed in almost exactly the same manner by both workers. For example, one subject drew a chest over the male figure pattern (Fig. 4). On inquiry, he described the chest as being "locked, but still a nice piece of furniture." One interpretation stated, "Subject has positive attitude toward father, but finds him inaccessible." The other said, "Subject is fond of father, but finds him hard to contact." No directly conflicting analyses were obtained. Judgments in which different points appeared to be emphasized were never found to be actually incompatible. As might be expected, the writer's previous experience with such material enabled her to make twice as many interpretations as were mentioned by the other analyst. However, in 96 per cent of the interpretations made by the two analysts which referred to the same behavior, there was no substantial difference in the points they had made. This agreement probably indicates the value of the inquiry statements used in guiding these analyses.

The second check dealt with the problem of whether the analysis of a record was modified by the analyst forming notions about the case from the first few responses of a subject, and subsequently making interpretations on the rest of the subject's responses in terms of these preconceived notions. Another person was asked to select at random a set of eleven subject responses to the stimulus-patterns from the complete materials at hand. This set of eleven subject responses was thus a composite record, in which each subject response was taken from a different case. Then the writer analyzed this composite as though it were a single subject's responses. In the analysis of the composite record, one interpretation was made about each subject response. The interpretations were compared with those originally made on the individual subject responses when they had been considered in their context. All eleven interpretations made in the check had been given in the original analyses also. Considering the eleven stimulus-responses, what seemed to be the central concept in each subject response formed the interpretation. All eleven interpretations proved to have the same central concepts previously connected with these subject responses. This check is possibly somewhat inconclusive because the writer was aware that the set-up was artificial. Nevertheless, this procedure has made it seem unlikely that the analytic process is much affected by preconceptions formed in advance of the author's analysis of each case.

CHAPTER III

SAMPLE ANALYSES

SOURCE OF MATERIAL

In order to make clear just how the various principles discussed in Chapter II were actually applied, two records, completely analyzed, are presented with a step-by-step explanation of the method used. *Case A* is taken from the Normal Children Group, in which validation was by clinical conference records, including data from Rorschach, TAT, interviews, etc., as listed in full above (see Chapter II). No inquiry material was obtained in this group, except for the spontaneously given labels supplied by some subjects, as in this case. The reasoning used in arriving at interpretations is indicated. The quotations from the clinical conference record relating to the interpretations are fully presented. *Case B* is from the personal interview group. The complete inquiry is quoted with explanations of how this material was used in forming the interpretations. A verbatim account of the interview section is included.

Case A: White Boy, Age 13

The following case is an example of procedure in the blind analysis of a case in the Normal Children Group. It will be recalled that in this group no inquiry was administered. The material used for validation has been more fully described above (see Chapter II). It is chiefly a summary of independent analyses by other workers using several different instruments such as TAT, Rorschach, interview, self-rating questionnaires, etc.

Below we present the analyses that were made, the rationale for the analyses, direct quotations (from the validating material) of such statements as seem to be pertinent and the scoring category into which the interpretation was placed. It is important to keep in mind the fact that no verbal associations were available in analyzing this record.

SYMBOL I (Fig. 12).

Analysis:

Although without inquiry it is difficult to tell exactly what the effect connected with the drawing might be, it is assumed that the sun is probably a source of warmth and at the same time, a distant object so far as the subject is concerned.

Two interpretations were formed here:

A. Mother fixated.

B. But the mother seems remote to subject.

Validation:

A. Mother tie largely a fixation. (TAT)

B. . . . Marked by not over-affectionate protection.

Early lack of contact with mother. (TAT)

Uncertain of mother. (Rorschach)

Scoring:

Since the above quotations from the validation material seem to be practically a paraphrase of what was stated in the blind analysis of the drawing material, both IA and IB were scored plus.

SYMBOL II (Fig. 13).

Analysis:

It seems to be a safe guess that the attitude toward a "lollypop" suggested a positive approach. The oral element seemed obvious, too. For these reasons the following analysis was made:

A. Oral fixation.

B. Positive attitude in individual heterosexual situation.

Validation:

A. Mother dependence and unfulfilled succorance needs. (Rorschach)

B. Positive interest in heterosexual situation. (Rorschach)

Sexual interests that are developing. (TAT)

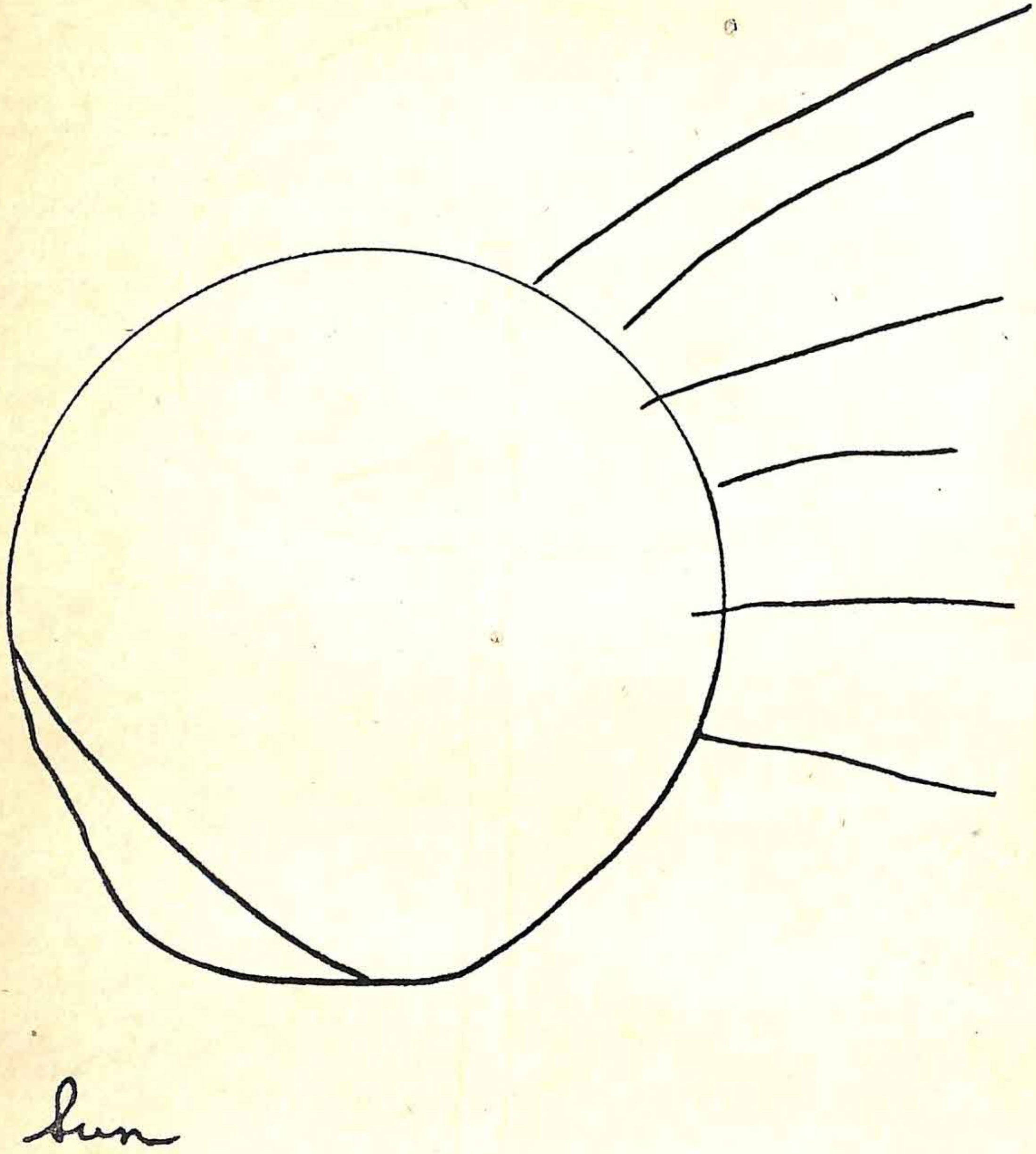


FIG. 12. Case A: Symbol I.

Scoring:

"Unfulfilled succorance"[1] seems to imply early deprivation in the context of mother dependency. For this reason it was felt that the statement from the validation material supported the likelihood of an oral fixation in this subject, although it hardly proved the matter. A "partial-plus" was the scoring used here.

The hypothesis that there was a positive attitude toward heterosexual situations seemed fully supported by the validation material, and the scoring for this hypothesis was plus.

SYMBOL III. (Fig. 14).

Analysis:

The fact that the subject drew a good figure out of the symbol without changing its essential

---

[1] Term used in clinical conference to describe desire for passive role in which needs are supplied by an outside source.

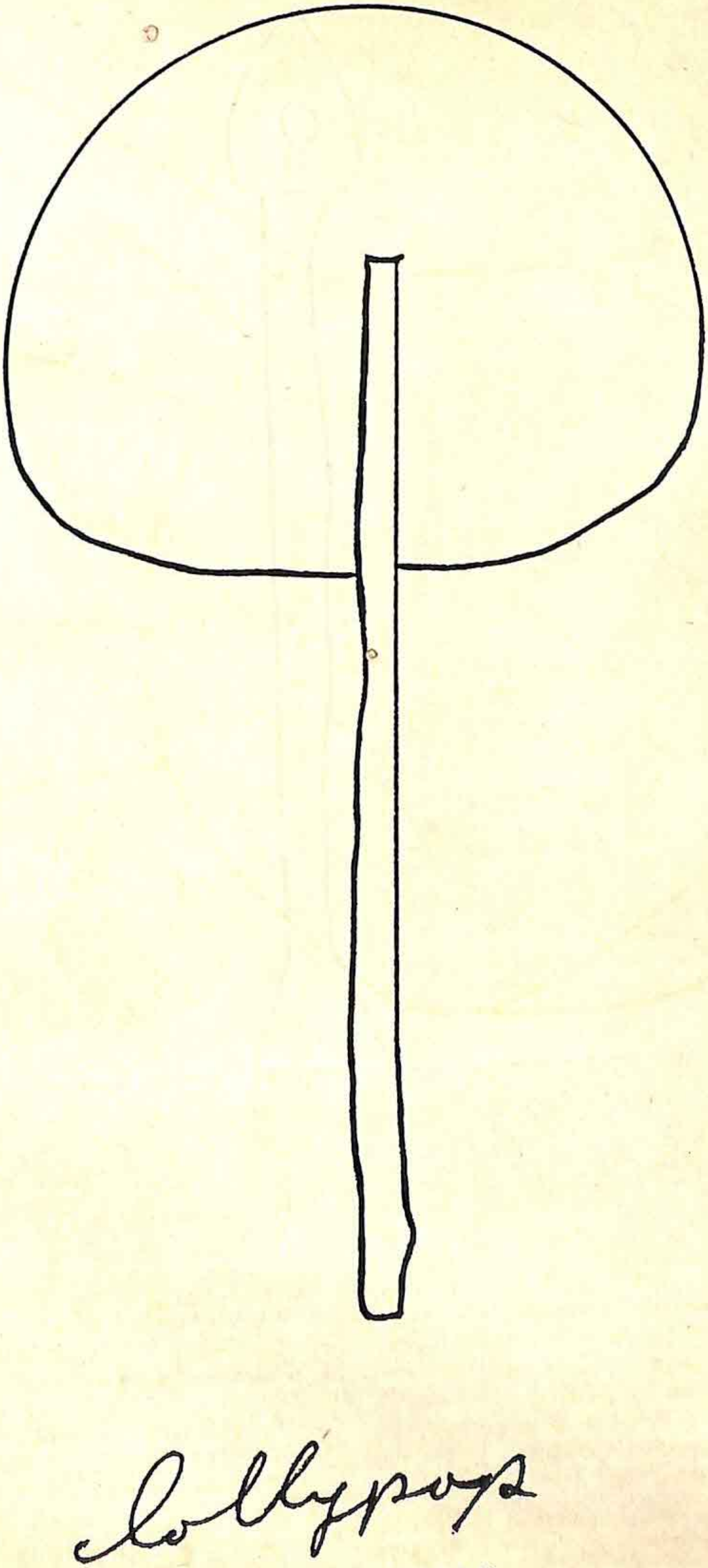


FIG. 13. Case A: Symbol II.

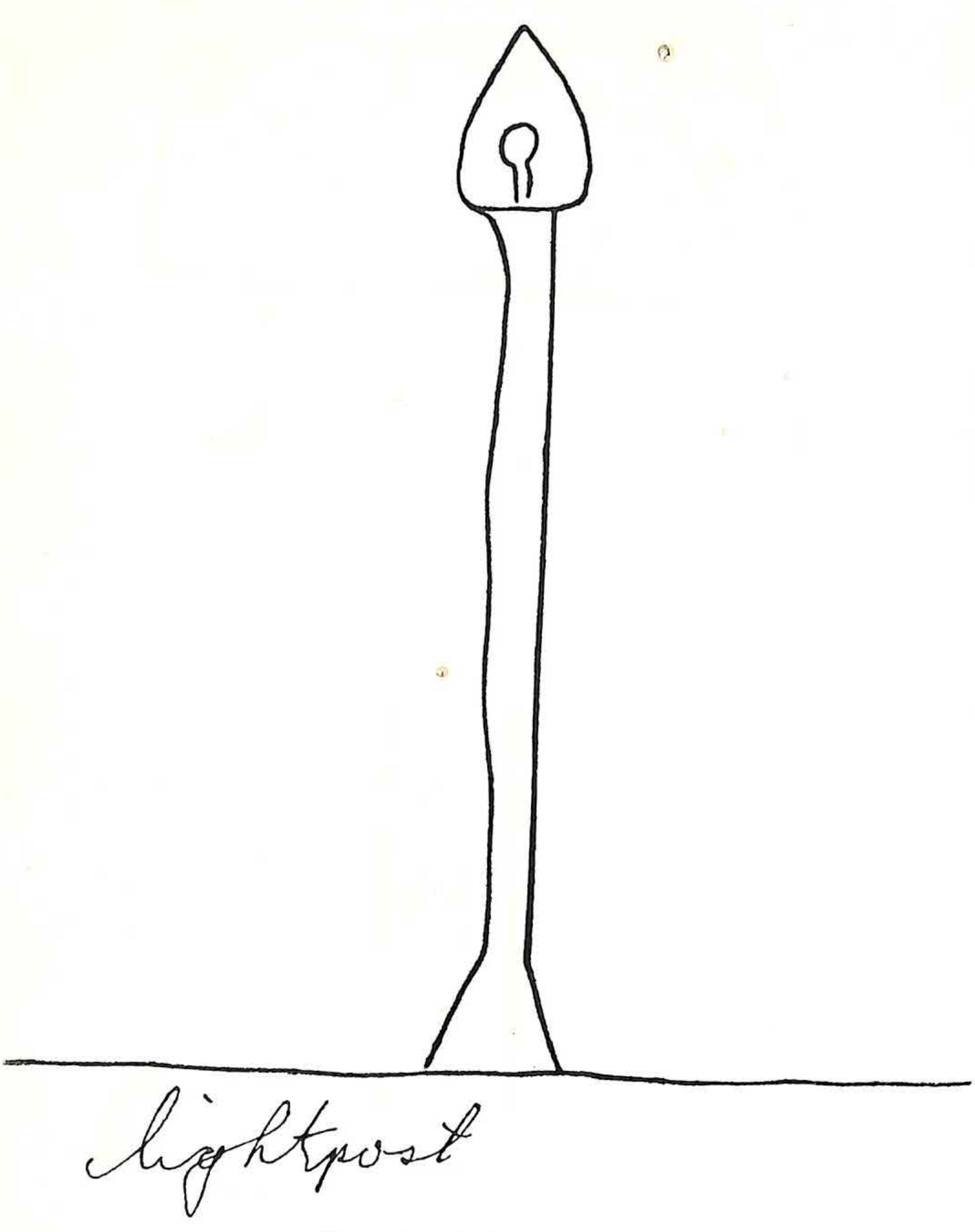


FIG. 14. Case A: Symbol III.

form seemed to indicate an acceptance of his father's masculinity on a positive basis. The fact that a lightpost was chosen would seem to indicate that the boy tried to associate guidance with the father figure. The analysis was written up as follows:

A. Tries to identify with father.
B. Father is the source of standards.

Validation:

A. Father first person mentioned on list of "people I'd like to be like."
Strives to identify with father. (TAT)
B. Father source of strict, uncompromising moral code to which subject shows resistance but probably will internalize. (Rorschach)

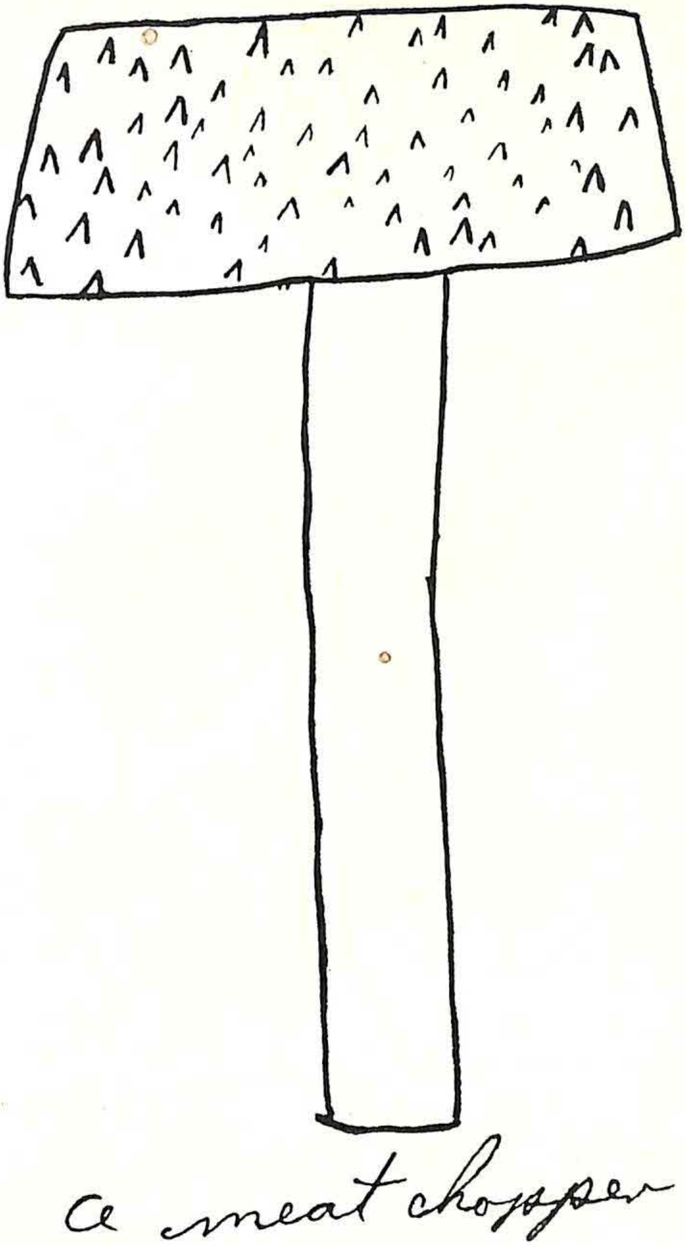


Fig. 15. Case A: Symbol IV.

Father man of firm demands which subject tries to obey. (TAT)

Scoring:

It seems entirely justified to assume that the validating statement completely corroborated the hypothesis. The scoring therefore was plus for both statements.

SYMBOL IV (Fig. 15).

Analysis:

It seems likely that the control of sexual anxiety is somehow connected with display of force. The following analysis was made.

A. Sexual anxiety controlled through display of force.

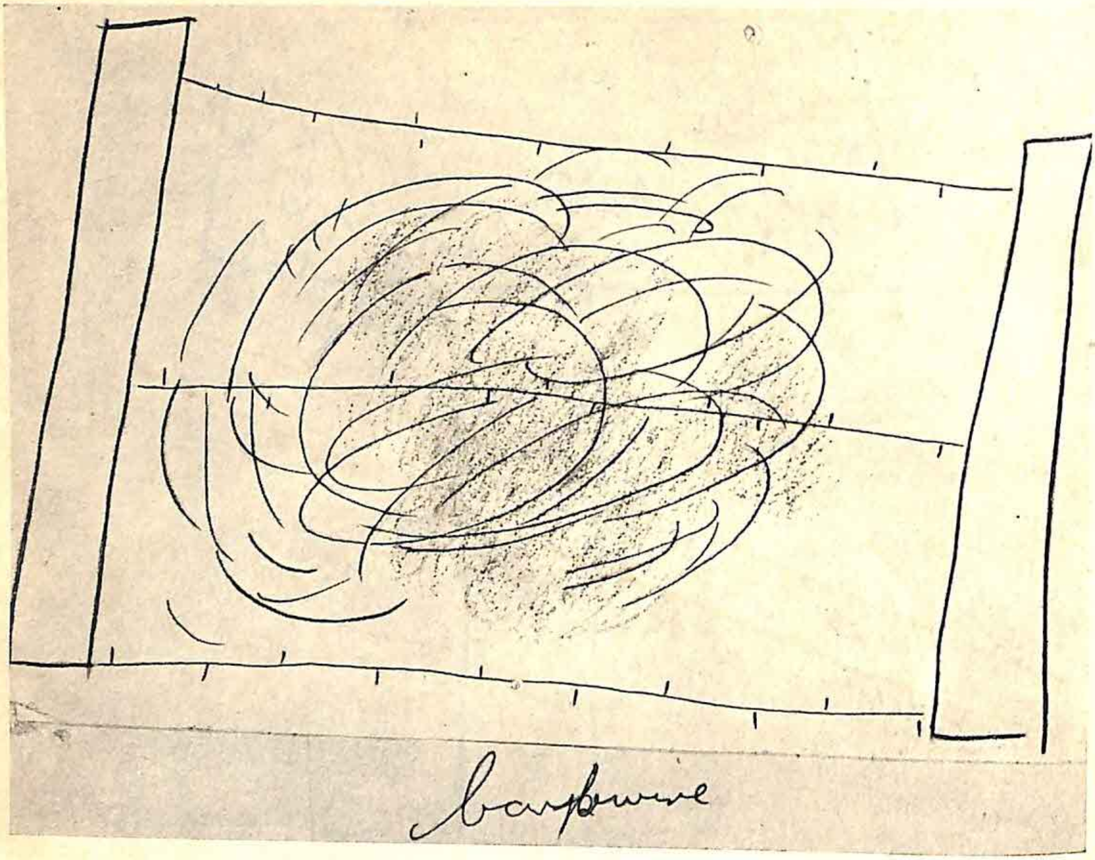


FIG. 16. Case A: Symbol V.

Validation:

A. Sent anonymous notes to older woman demanding she meet him in shack. (Case history) Insisted upon identification that no harm was meant. (Case history)

Scoring:

Because the only evidence in the record is but remotely related to the interpretation, and it is conflictual, the scoring was "partial-plus-minus ($P\pm$)."

SYMBOL V *(FIG. 16).*

Analysis:

It did not seem far-fetched to assume that a barbed wire fence is connected with environmental restriction. Analysis was written as follows:

A. Much anxiety concerning environmental restrictions.

Validation:

A. First and foremost, subject struggled with emotional problem of dependence and independence. (Rorschach)

Anxiety chiefly in terms of negativism, more the situation-reactive kind, rather than some deep-seated disturbance . . . is source of guilt over resistance to parental control. (TAT)

Scoring:

The quotations from the validation material seemed in this instance to justify a full plus scoring for the hypothetical analysis.

SYMBOL VI (Fig. 17).

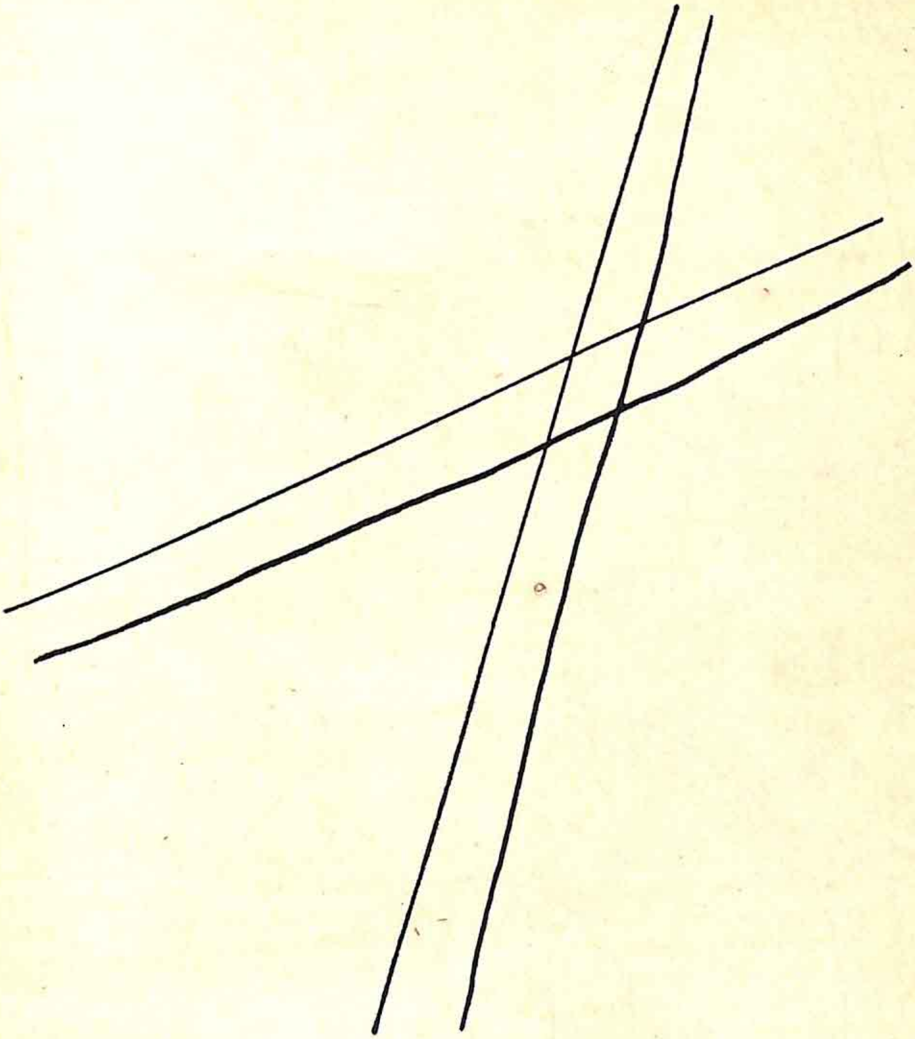
Analysis:

The only thing of which one can be sure, and then only relatively so, in this case, is that there is a transitory quality expressed in the drawing. For this reason, the following analysis was made:

A. Finds only transitory contact and no deep relations possible with other boys.

Validation:

A. Negative feeling of group toward him . . . he makes aggressive attempts to join groups. (Sociometric analysis)



two highways crossing on a angle

FIG. 17. Case A: Symbol VI.

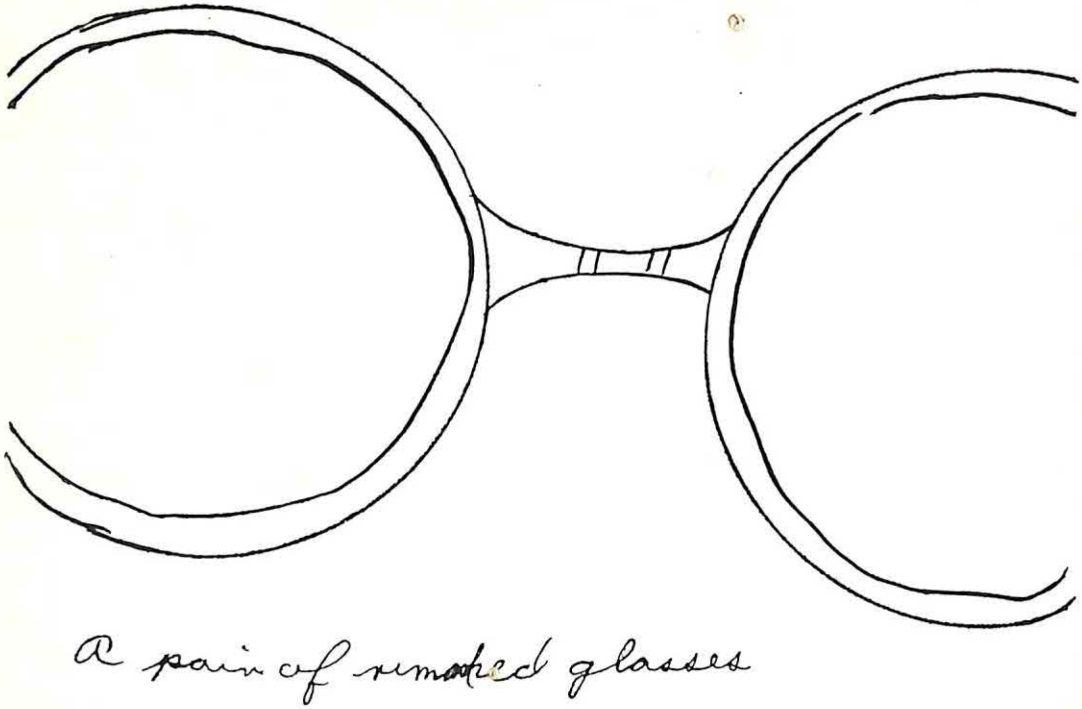

a pair of rimmed glasses

FIG. 18. Case A: Symbol VII.

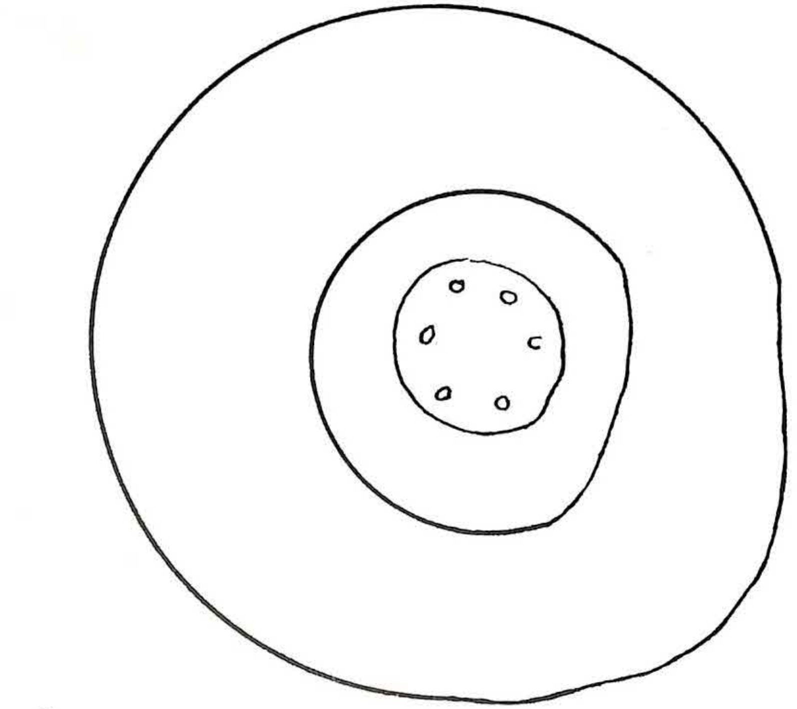

a car wheel and tire

FIG. 19. Case A: Symbol VIII.

Scoring:

The evidence in this case seem to warrant a full plus scoring.

SYMBOL VII (Fig. 18).

Analysis:

Without having more information from the subject, one might connect glasses with a wish for "better vision," and thus with a desire for better understanding. (We often say: "I see" when we mean: "I understand".) There is also a need for shielding or protection symbolized by glasses, and thus we hypothecate a withdrawal reaction too. The statements made in the analysis of this drawing were as follows:

A. Striving for greater understanding in orientation to opposite sex.

B. Also has defensive withdrawal reaction with girls.

Validation:

No statements were found which related to the subject's attitude toward women in general or, particularly, the female members of his peer group.

Scoring:

The scoring here was *"O."*

SYMBOL VIII (Fig. 19).

Analysis:

Closing the figure in this symbol has seemed, in our experience, to indicate a regressive personality trend. Also, though it is not clear in the reproduction, the subject wrote "care" at first, and then erased the "e." The slip made in the spelling of the word "car" seemed to justify hypothecating a dependent attitude on the part of the subject. The analysis was as follows:

A. Regressive and dependent.

Validation:

A. Strong childlike emotional dependency. (Rorschach

Scoring:

The quotation from the validation material seemed to furnish corroboration of the interpretation. The scoring was "plus."

SYMBOL IX (Fig. 20).

Analysis:

From this drawing, even though it was labeled as it was, no analysis seemed possible without further information, except as a complete guess. Therefore, no analysis was attempted.

SYMBOL X (Fig. 21).

Analysis:

Without further information, it seemed that only a statement of the passivity of the subject's personality could be made. A relationship has been noted between the drawing of vessels of various sorts on this symbol, and a passive attitude. The following statement was made about this symbol:

A. Passive personality.

Validation:

A. Aggressive, argumentative, although peers feels his aggressiveness is inconsistent. (Case history and sociometric material)

B. Emotionally dependent. (Rorschach)

Scoring:

There does not seem to be a statement directly bearing on the question of passivity, and so both of the above quoted statements must be accepted, even though they are contradictory. The scoring was "plus-minus (±)."

SYMBOL XI (Fig. 22).

Analysis:

The general crudeness of the drawing plus the resistive implications of "crossed legs" seem to warrant the following analysis:

A. Feels incompetent in competitive heterosexual situations.

B. Feigns indifference to such situations.

Validation:

No evidence to be found bearing on the subject's relationships in the competitive heterosexual situation.

This case illustrates analysis by content alone. The contrast with the following case makes it clear how useful the verbal association material is in gaining understanding of the subject's drawing responses.

Case B: White Female, Age 24

This subject was a member of the Normal Adult Subgroup B (interview group). She had been married for five years. When she volunteered for this study, her companion stated to the writer, "You now have in your group that rare thing, a really model wife." It might also be added that she was a person of very superior intelligence, a leader-type who had proved her ability in a

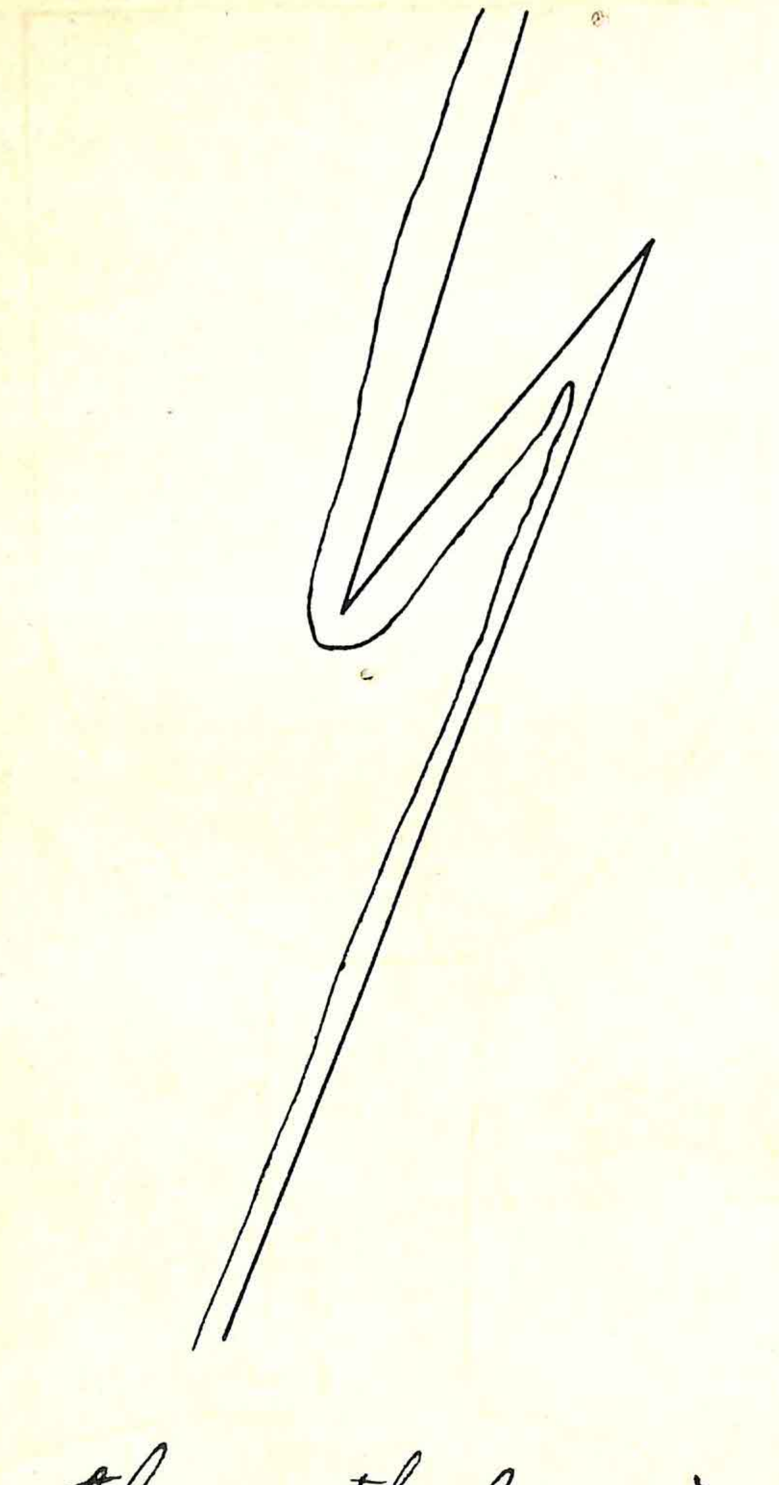

the path of a river.

FIG. 20. Case A: Symbol IX.

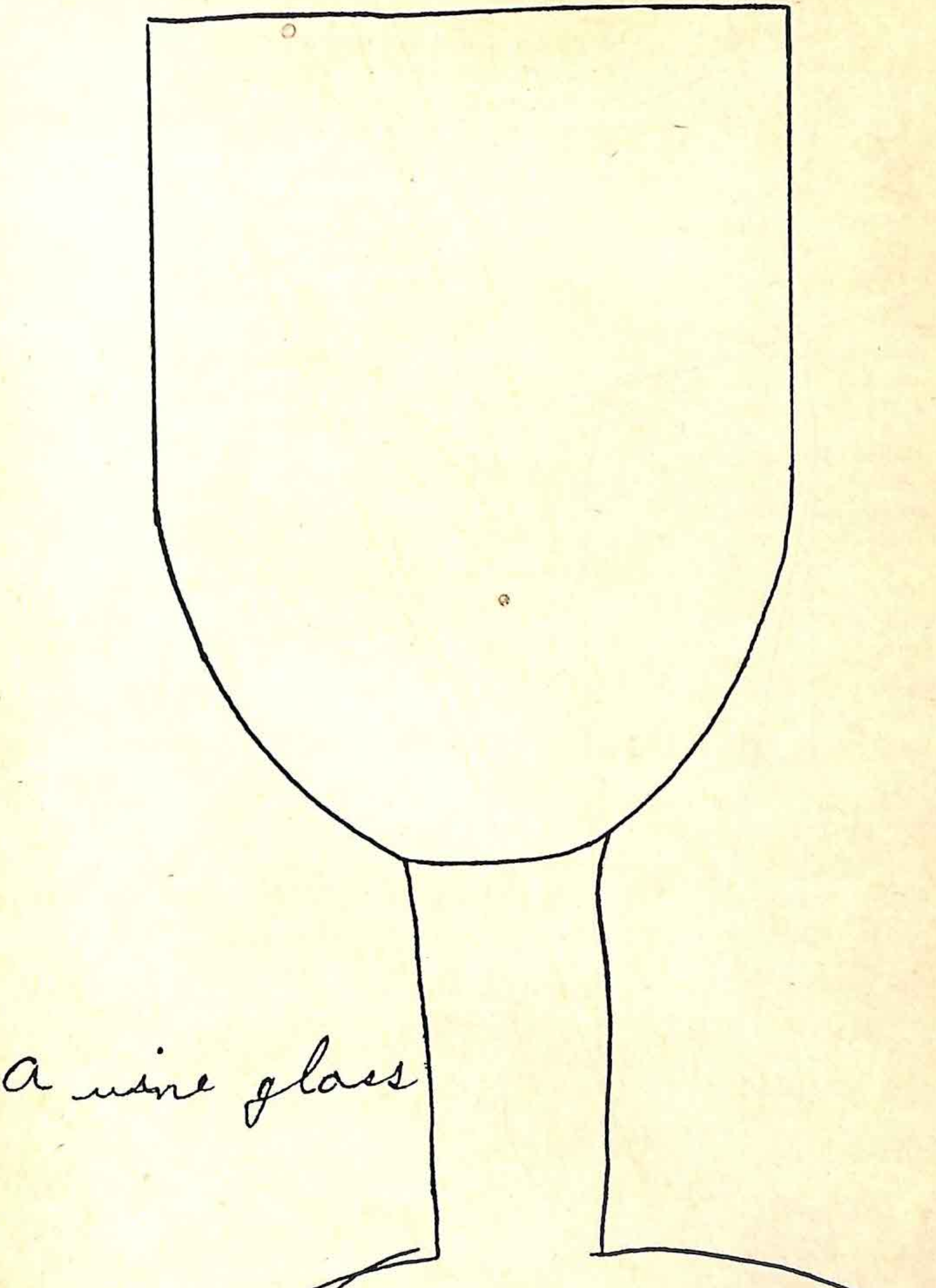
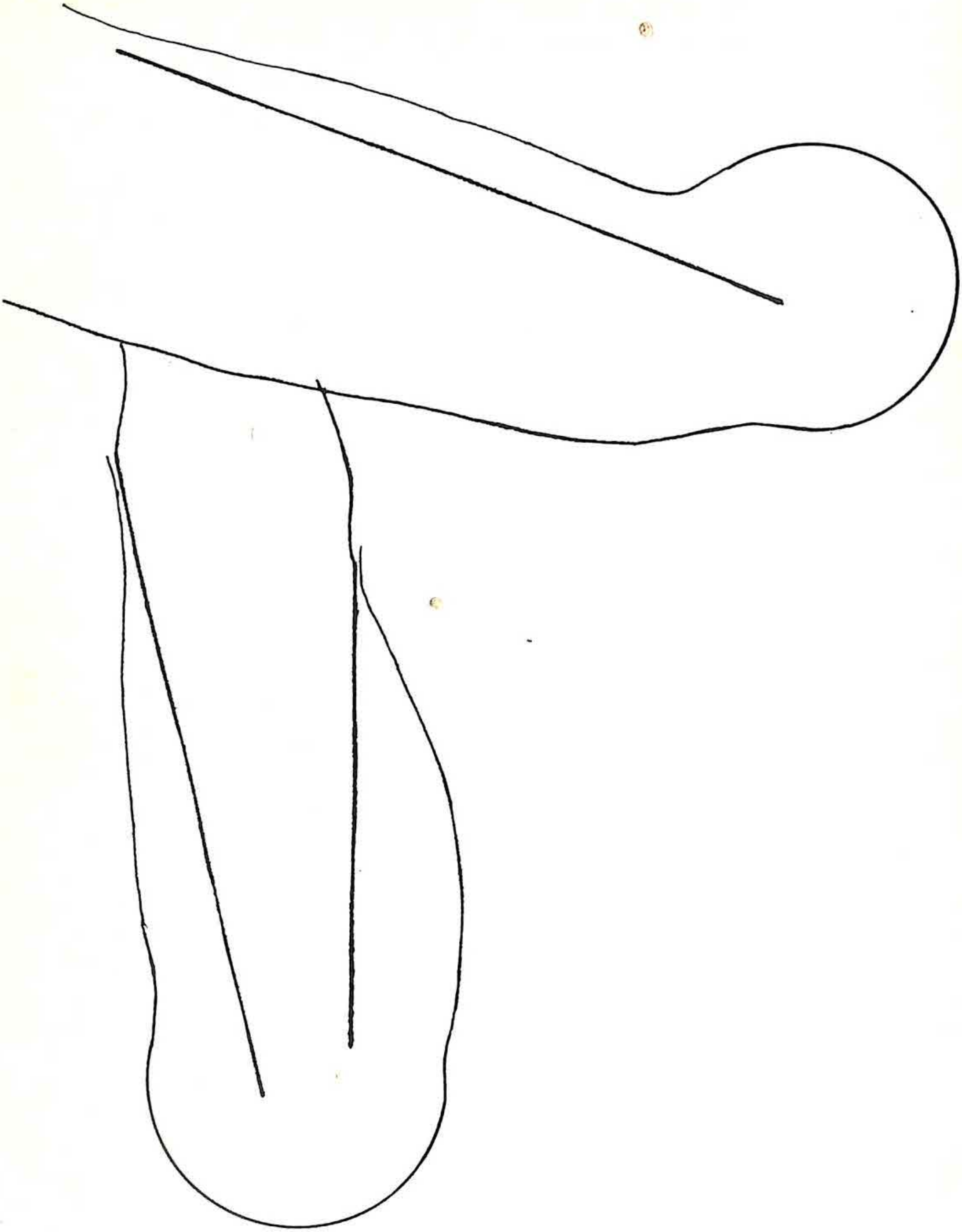


FIG. 21. Case A: Symbol X.



A person sitting there legs crossed.

FIG. 22. Case A: Symbol XI.

number of ways. This is especially important because some of the questions asked might at times appear like leading questions. Knowing this personality is to realize that she was not suggestible, and possessed the facility to see quickly through and into a question asked.

The subject was given the directions described in Chapter II. She finished the drawings unusually quickly, taking little more than a half an hour for the entire performance. Then she was given the inquiry questions which she proceeded to fill out in the way in which they appear on the copies of the drawings. The following pages describe the process of analysis in detail.

Included in the interpretations that follow is a verbatim account of the interview with the subject.

Following is the list of inquiry questions; these may be used as a guide in interpreting the written remarks of the subject, which were recorded after she had completed each drawing.

A. What is this a drawing of?
B. Was it relatively easy or hard to draw?
C. What mood does it represent?
D. Describe in detail the personalities of any people or animals depicted:
  1. What are they doing?
  2. What are they thinking or feeling?
E. What mood do you feel as you look at it?
F. Do you like it?
G. What does the drawing remind you of?

Interpretation is based upon the answers to these questions.

SYMBOL I. Semi-circle Pattern (Fig. 23).

Interpretation 1:

The fact that all the answers to the inquiry questions expressed positive affect was interpreted as meaning a positive attitude toward the mother.

Q. "Are you particularly fond of your mother, perhaps closer to her than to your father?"

A. "Oh, yes!"

This emphatic reply seemed to be sufficient for validation purposes.

Interpretation 2:

Since the subject labelled the structure drawn from stimulus-pattern "A library," it was assumed that there might be an interest in intellectual pursuits which was stimulated by the mother. This likelihood, of course, depended on the accuracy of the first interpretation that the subject had an affectionate regard for her mother.

Q. "Do you share interests with either parent particularly?"

A. "Well. My father is as intellectual as my mother . . . he is a teacher and writer, but my own interests are definitely like my mother's, especially in politics. She always used to guide my reading."

The fact that the subject took pains to establish that she might have taken her intellectual cues from her father but did not is evident. It is interesting that she placed her answer in an intellectual context also, even though the questioner had expected to have to ask a second question in order to validate her interpretation.

Interpretation 3:

The fact that the subject reports her own mood as being "nostalgia" when looking at the drawing suggests the interpretation that she and her mother are not as close as they once were.

Q. "Have you and your mother been as close lately as you once were?"

A. "Oh, no. We still live next door to each other, but for the past year or so—since she has opened her shop—I hardly see her at all."

The subject's tone as she said the above was somewhat depressed, it seemed to the questioner. In any case, it was clear from her answer that the facts validated the interpretation.

SYMBOL II. The Semi-circle and Single Line Pattern (Fig. 24).

Interpretation 1:

The inquiry indicates positive affect. It is also noteworthy, however, that the mood of the picture is called "relaxation," and the subject says that "everything is taking it easy." The interpretation made was that the subject had a positive but passive attitude toward intimate heterosexual relations. Since she is married, it was assumed that this attitude was reflected in her relations with her husband.

Q. "How would you characterize your relations with your husband?"

1. A public building.
2. Easy.
3. Exhilaration.
4. No particular personality.
5. Nostalgia.
6. Yes.
7. A library.

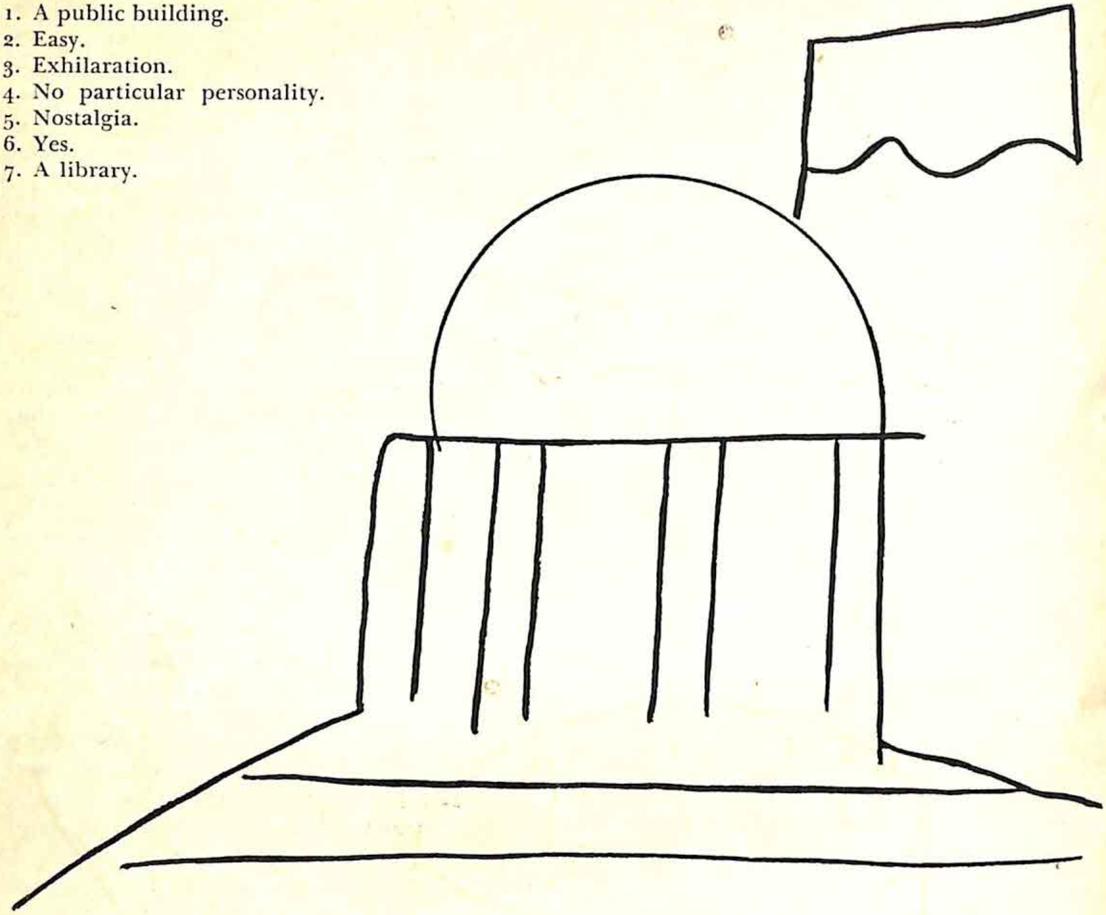


FIG. 23. Case B: Symbol I.

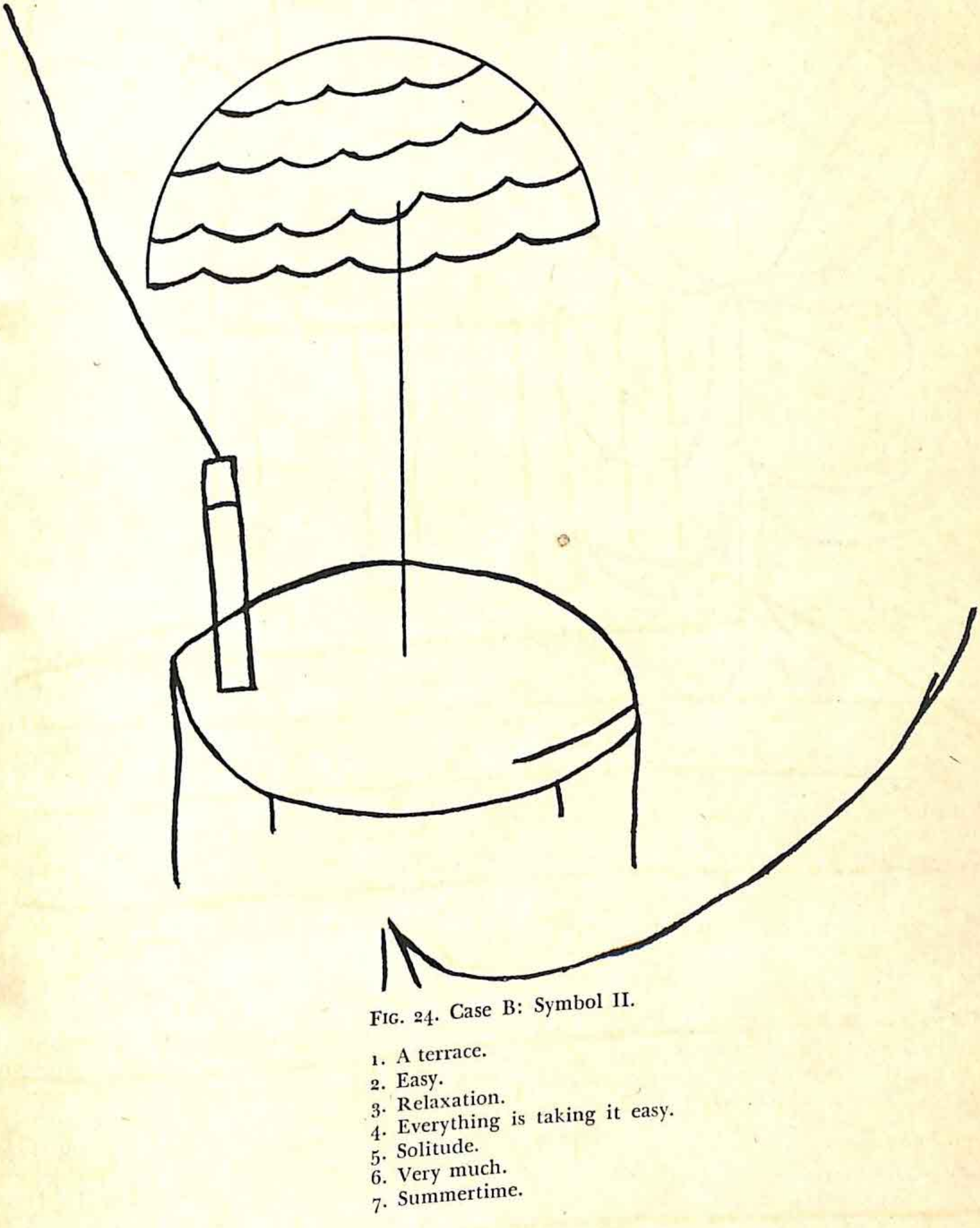


FIG. 24. Case B: Symbol II.

1. A terrace.
2. Easy.
3. Relaxation.
4. Everything is taking it easy.
5. Solitude.
6. Very much.
7. Summertime.

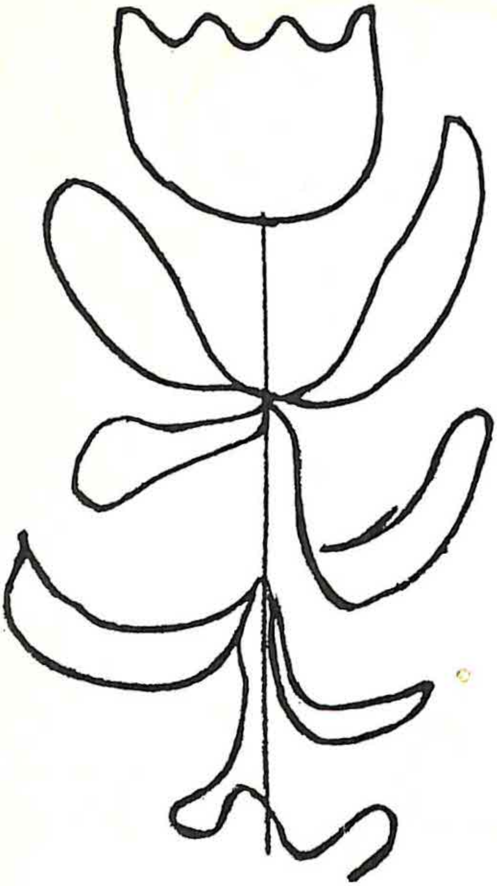


FIG. 25. Case B: Symbol III.

1. A tulip.
2. Easy.
3. Gaiety.
4. Restraint (personality of the flower).
5. Gaiety.
6. No.
7. Amy Lowell's "Patterns."

A. "Oh, very pleasant."

At this point it seemed that a more direct form of question could not be avoided.

Q. "Your relations with your husband are satisfying, but more characterized by a feeling of tenderness perhaps rather than 'purple passion'?"

Q. "Yes. How did you get that?"

A. "The object of this research is to test whether this kind of material can be used in this way."

SYMBOL III. The Single Straight Line Pattern (Fig. 25).

Interpretation 1:

From the fact that the subject expressed liking on the first figure and negative affect on this one, it was deduced that the subject's relations were less adequate with her father than with her mother.

Q. "Could you say which parent you prefer?"

A. Easily. It's a funny thing, but you see—my father was disappointed. He wanted a boy. I'm afraid that he never did come to feel very good about me. So I could hardly be as close to him as to Mom. Of course, things are pretty cheerful on the surface with my father, too."

This response seems to be acceptable as a confirmation of the interpretation offered.

Interpretation 2:

The fact that the content was a flower in this drawing seemed to be significant. Whether or not one might give gender to a flower[2] it seemed safe to consider such a drawing on the "maleness" stimulus as depicting a relatively "weak" figure.

Q. "Which parent would you say was the more forceful in your family?"

A. "My mother has always dominated the family."

This answer seems to supply sufficient confirmation for the interpretation.

Interpretation 3:

Since the "personality of the flower" was considered to be "restraint" (Question 4 on the inquiry), it was supposed that the same characteristic would be true of the personality of the subject's father.

Q. "Would you say that your father's attitude toward you was one of restraint?"

A. "I should say so!"

It is obvious that more might have been said in this analysis, particularly with regard to the reason for the mood of gaiety and the poem mentioned in Question 7 in the inquiry. However, as has been explained above, no more than three hypotheses were made apropos of any one symbol. Possible connections with husband were not discussed.

SYMBOL IV. The Inverted "V" Pattern (Fig. 26).

Interpretation 1:

Since all the associations on the inquiry seem

---

[2] In the vernacular of the street, a passive male homosexual is referred to as a "pansy."

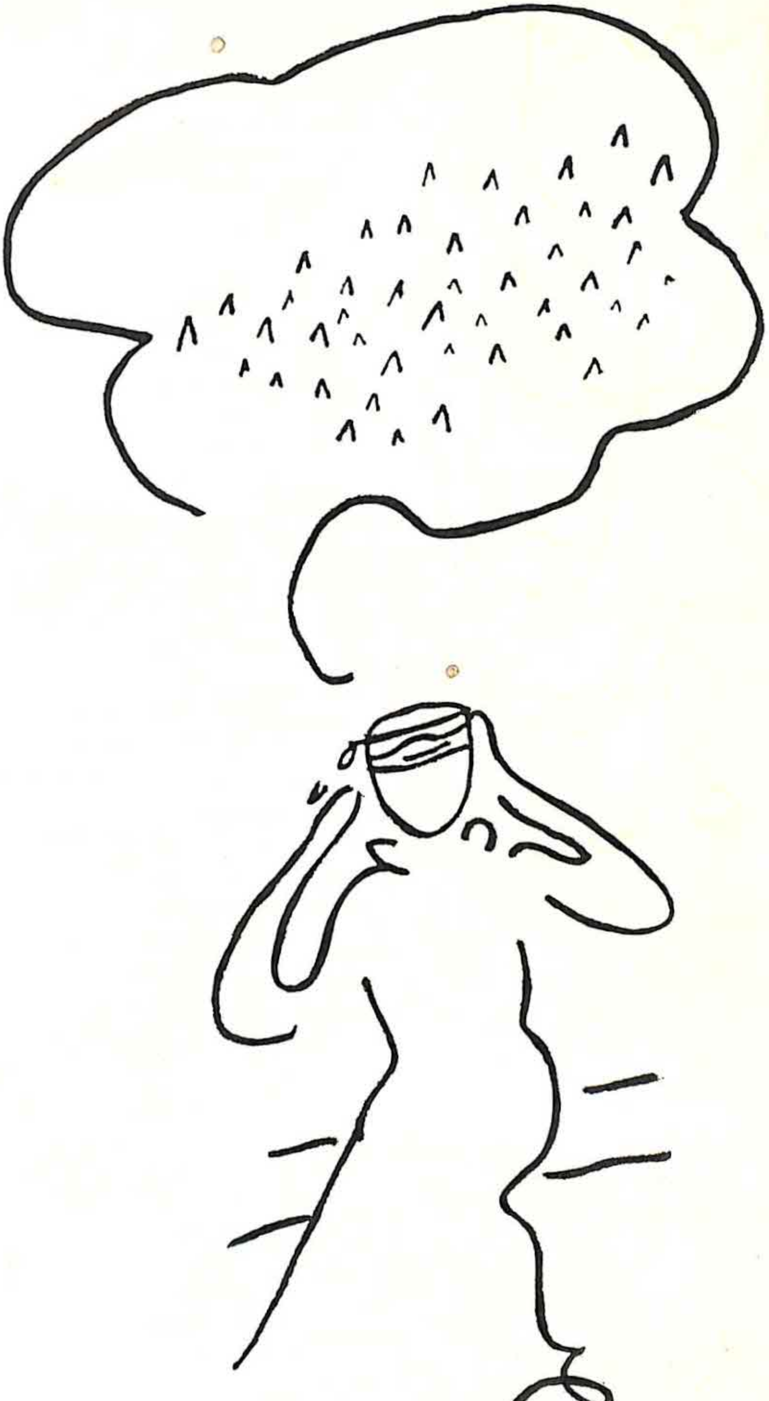


FIG. 26. Case B: Symbol IV.

1. Woman nursing a headache.
2. Hard.
3. Torment.
4. She is wearied with nursing grudges and occupies her time with trading neighborly pettinesses.
5. Anguish.
6. No.
7. My mother-in-law.

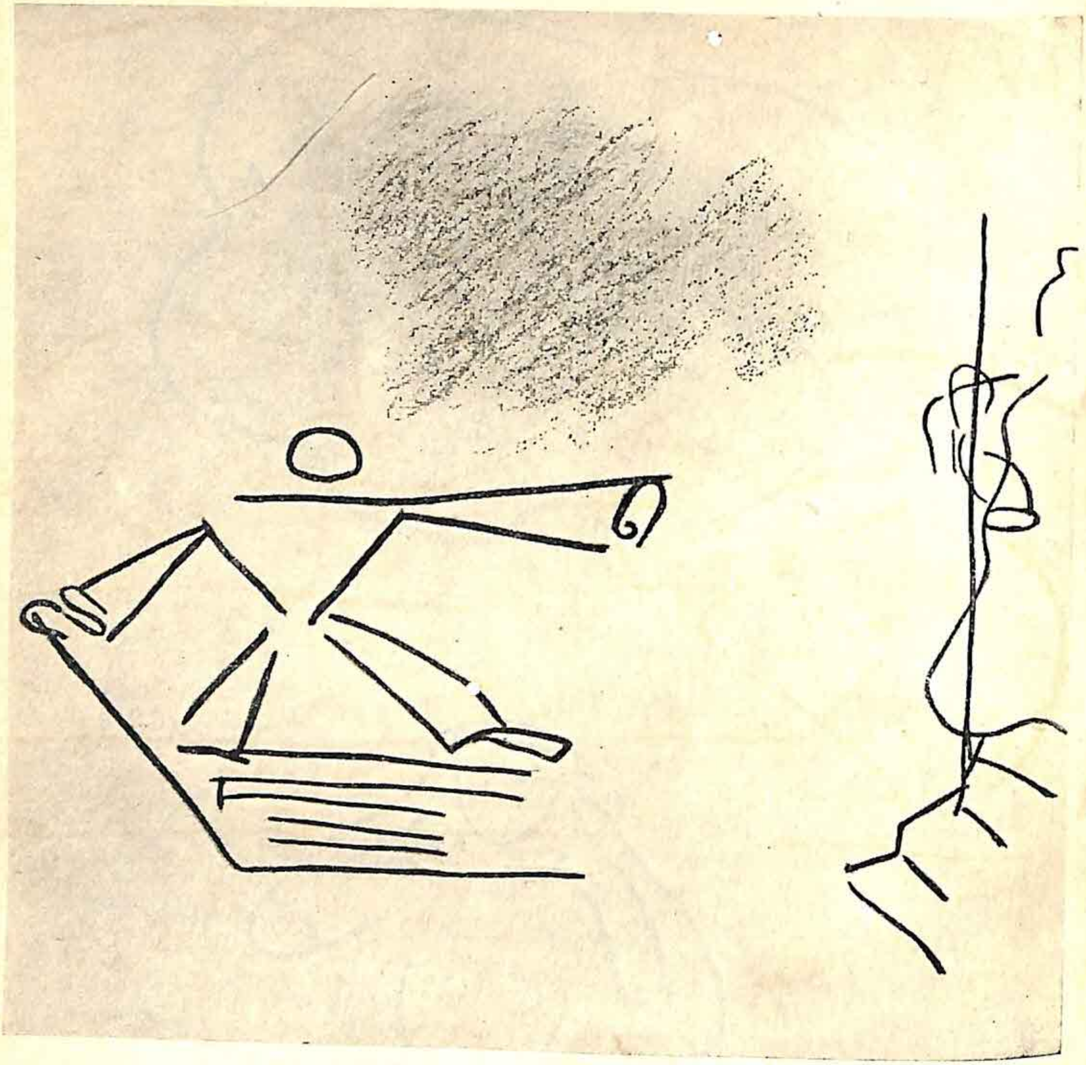


FIG. 27. Case B: Symbol V.

1. A tornado.
2. Easy.
3. Fear.
4. The man, like Per Hansen, is a stocky, broad-shouldered man; the woman is a stocky broad—he is concerned about getting his family to safety; she is wondering which of the household goods she should trundle along.
5. Determination.
6. Yes.
7. Kansas.

to refer to the subject's mother-in-law in a negative manner, it was assumed that there must have been conflict which affected the subject's sexual adjustment. The reaction to the symbol for sexual anxiety is not always interpreted in reference to intimate heterosexual relations. In this case, this seems to be the most likely interpretation.

Q. "You seemed somewhat wistful when you agreed that 'purple passion' does not characterize your relationship with your husband. To what do you attribute its absence?"

A. "I don't know . . . I guess the story-book phase of our honeymoon is over."

Q. "How has it changed?"

A. "Well, sometimes . . . I don't know, I get the

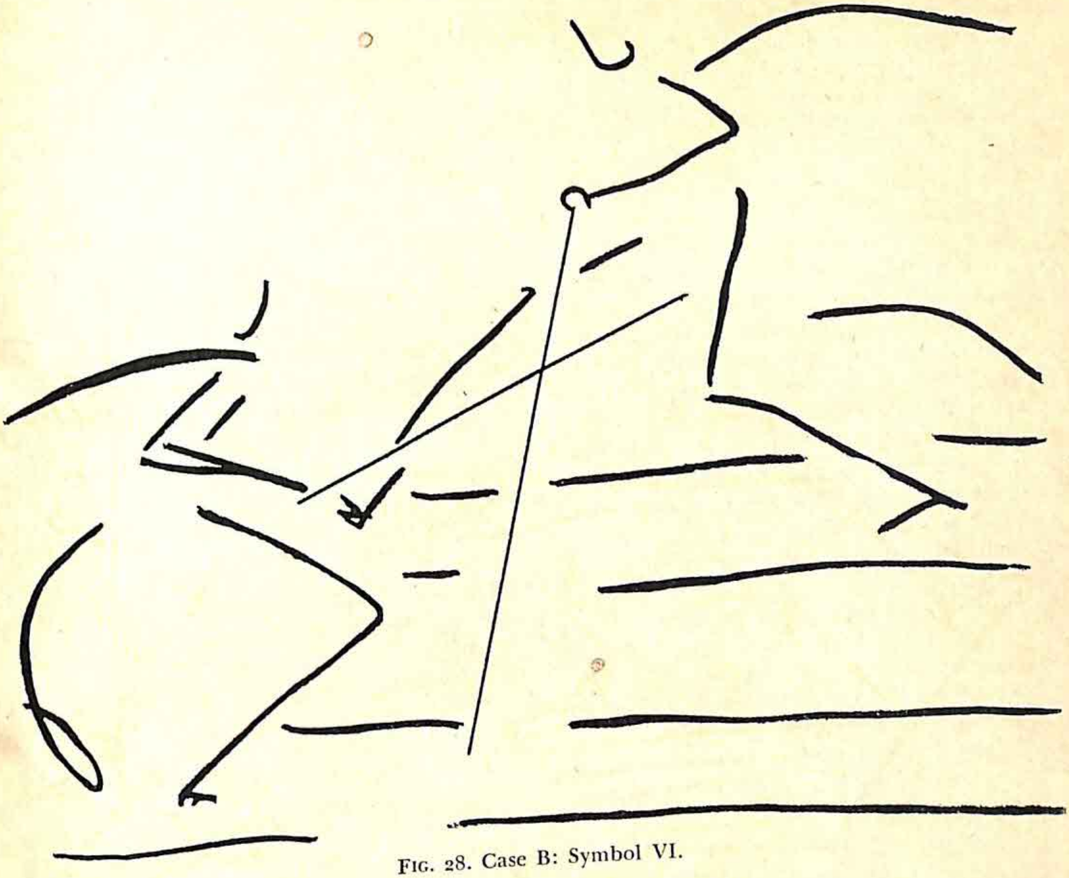


Fig. 28. Case B: Symbol VI.

Symbol No. 6. The Crosslines Pattern

1. Two fencers.
2. Easy.
3. Romance.
4. Stock characters out of a Basil Rathbone movie.
5. Excitement.
6. Yes.
7. The Beggars Opera.

feeling that he isn't all mine. *You'll* probably eat this up, but sometimes I feel there's too much mother-in-law. Now, don't get me wrong, I do like her. I mean she's all right—you know? It's just that she can't let him alone. Sometimes I think he's more tied to her than to me. I'm trying to change that, but he is having a hard time adjusting to being the loser. Oh, well, what else do you want to ask me?"

SYMBOL V. The Diffuse Mass Pattern (Fig. 27).

Interpretation 1:

The combination of the attitude of "determination" with the scene of strife suggested, in the context of the reactions previously ascribed to the man and to the woman, that the subject was determined to seek relief from the anxiety created by marital conflict.

Q. "Have you had serious concern for the future of your marriage, I mean, really, for a period of time?"

A. "Yes, when my husband returned from service I was quite sure that I could not change the plans that I had made to suit some of his new ideas. I don't know how he put up with me then. I used to pout all day long. However, there isn't any doubt about our staying together now."

Interpretation 2:

Since it was assumed that the people in the drawing were the subject's husband and herself, the interpretation was made that the characteristics mentioned in Question 4 of the inquiry applied to them.

Q. "How dependable would you say your husband is?"

A. "He's dependable if he's nothing else."

Q. "How would you compare your own reactions under stress with his?"

A. "Well, it's funny; he can be awfully silly. But when things get really tough, he's the one who considers the important values."

Q. "How do you mean?"

A. "Well, coming back to our marriage, he's the one who kept us together when superficial concerns almost made me run away. God! But I was miserable!

Well, let's go on to the next question."

Interpretation 3:

Continuing the logic of Interpretation 2, it was assumed that the subject considered herself to be materialistic, wanton.

Q. "Of course. But I'm afraid I must ask you another question on this topic. Which values do you feel you tend to emphasize in contrast to your husband, and is this anything that bothers you?"

A. "Well, he's not materialistic like I am. I *need* pretty things, for example, but he doesn't understand it. It still worries me at times but what can I do about it? He's right, I suppose, but I'm the way I am."

Q. "Sometimes you fear you are very materialistic and—um—somewhat wanton?"

A. "Yes, I'm afraid so. There isn't much doubt about that."

The last interpretation, although accepted by the subject, was not suggested to her in a way which gave her much freedom for discussion. It was considered to be only a "partial-plus" validation. However, the importance of her feeling that she tended to be materialistic in an anxiety-provoking situation is clearly confirmed.

SYMBOL VI. The Male Group Symbol (Fig. 28).

Interpretation 1:

It seems practically a paraphrase rather than an interpretation of the inquiry statements to say that the subject sees men in terms of romance and excitement, particularly in terms of conflict over her.

Q. "How do you feel about the idea of men fighting over you?"

A. "That's something any girl would enjoy, isn't it?"

Q. "Well, of course, now that you're married and 'out of bounds,' so to speak, there probably isn't much of a likelihood that men are actually fighting over you?"

A. "Oh, I would hardly say that they fight over me. But I'm not exactly a wall-flower either. Like at parties, I know that men still try to outdo each other to get my attention. That's what makes parties fun, don't you think?"

It may seem "natural," offhand, to feel as the subject does. However, many women do not feel similarly and some even disapprove of such an attitude. Thus, this interpretation is adequately specific for a particular subject.

SYMBOL VII. The Female Group Symbol (Fig. 29).

Interpretation 1:

The positive affect expressed on the inquiry suggested that the subject took a generally positive attitude toward other women.

Q. "How do you get along with other women usually?"

A. "Oh, I certainly enjoy working with the girls in the office and, of course, my best friends have always been girls. I know some women just don't like other women, but I can't see their point."

Interpretation 2:

The characterization in Question 4 of the inquiry suggested that the subject had a tendency to feel superior to most women, her positive attitude notwithstanding.

Q. "What do you think about women as a group, I mean in the sense of their being alert, and so forth?"

A. "It isn't that men are really superior, but most women get pretty stodgy—in a pleasant sort of way. But that's not because they *couldn't* be otherwise. I certainly wouldn't let myself go the way some woman can, and seem to enjoy it."

Interpretation 3:

The fact that the description in Question 4 of the inquiry seems to set the subject apart from the rest of the female group lent itself further to the interpretation that the subject had a resistance to pregnancy based on her unwillingness to identify with the female group.

Q. "How do you feel about having children?"

A. "Oh, I'm strictly a career-girl type, I'm afraid: No part of housewifery appeals to me. This 'babies and bridge-game' life most of

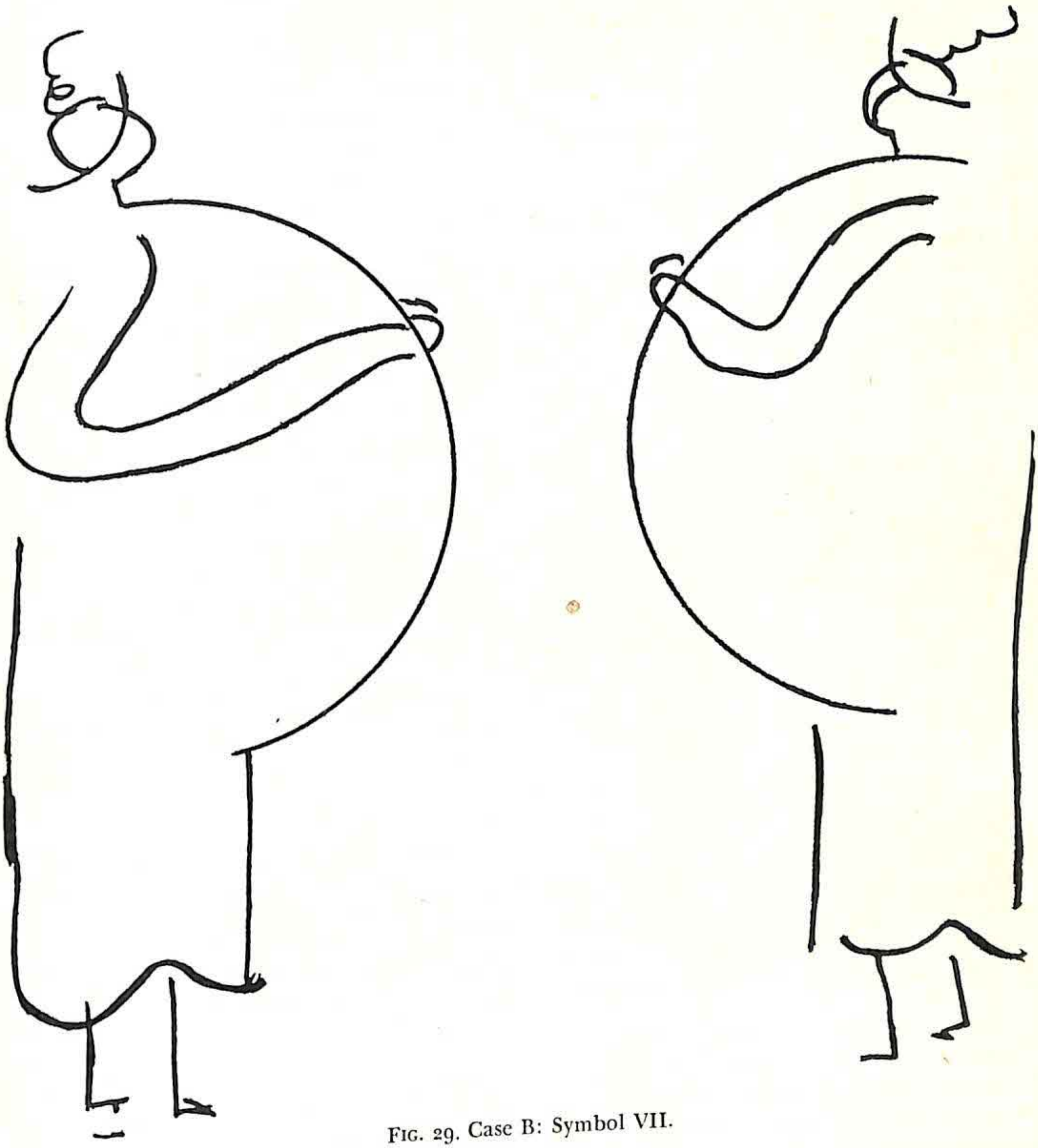


FIG. 29. Case B: Symbol VII.

1. Two pregnant women.
2. Easy.
3. Placidity.
4. Naive rather conventional women, fastidious enough to keep their skirt hems even despite their figure, very happy to do just as Jesus tells them to.
5. Happiness.
6. Yes.
7. Helen Hokinson.

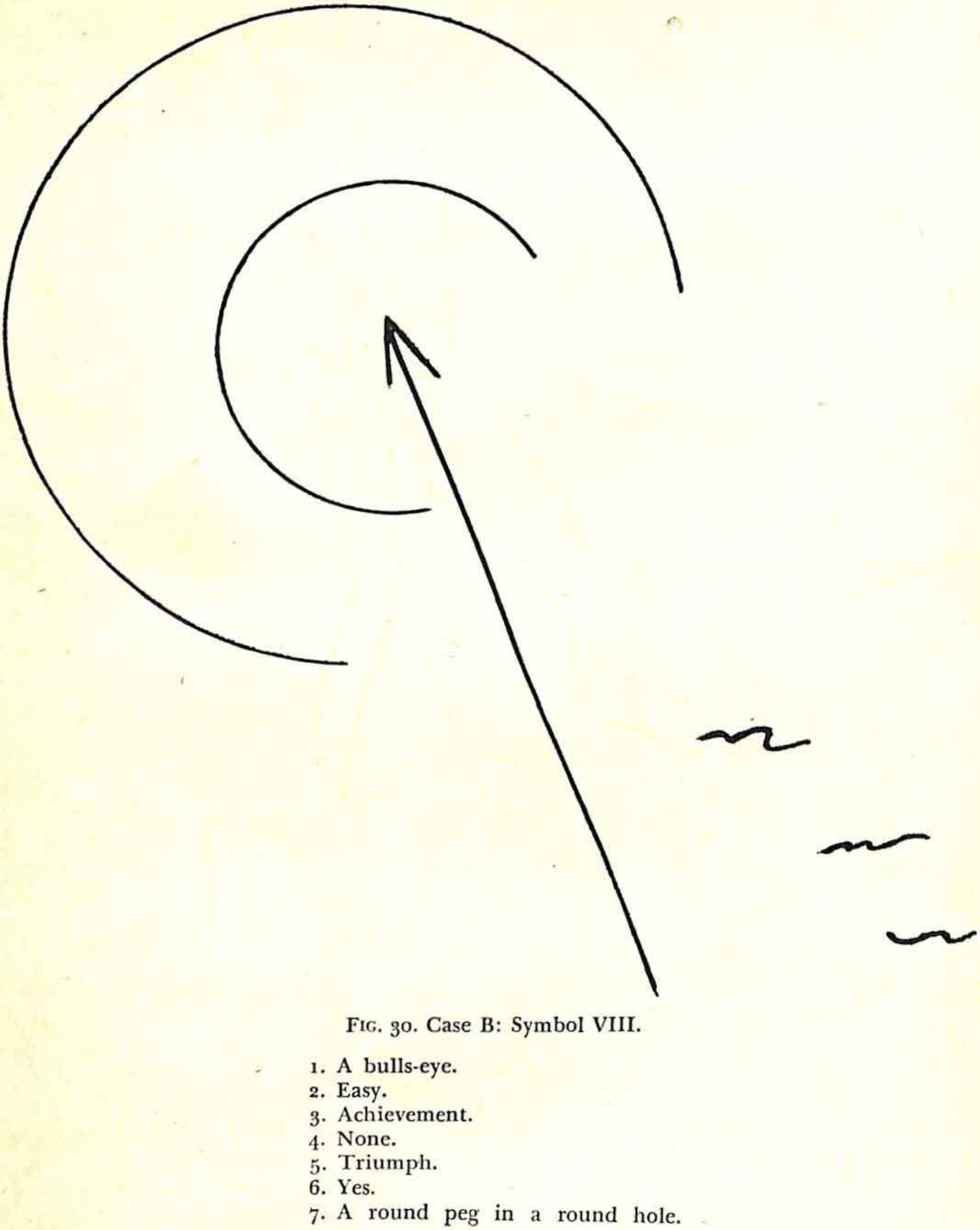


Fig. 30. Case B: Symbol VIII.

1. A bulls-eye.
2. Easy.
3. Achievement.
4. None.
5. Triumph.
6. Yes.
7. A round peg in a round hole.

my girl friends have compromised on is not for me. And the way some men like to tie their women down with babies! I'll have a baby when the man can do the suffering for me."

SYMBOL VIII. The Regression Symbol (Fig. 30). Interpretation 1:

The fact that return to the inner circle is characterized as being a source of triumph, and is associated with reaching a goal in the inquiry, seemed to indicate that the subject is very much security-conscious.

Q. "How important is security to you?"

A. "As a matter of fact, I've turned down a couple of agency offers quite recently, even though they would pay so much better, because I prefer staying at S——where I know

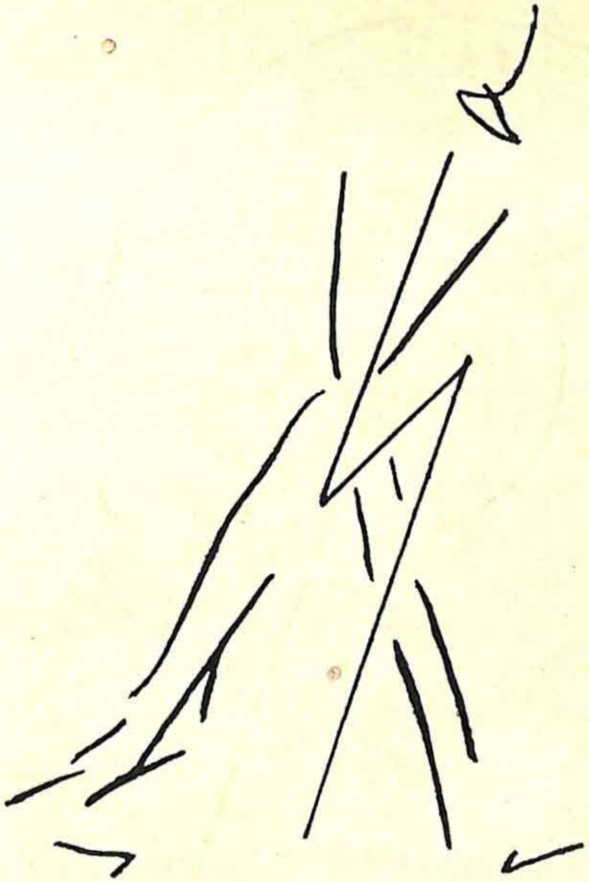


FIG. 31. Case B: Symbol IX.

1. Man walking, with a cane.
2. Hard.
3. Doggedness.
4. A man of taste, set in his own ways, indifferent to criticism, loving cities.
5. Stubbornness.
6. No.
7. A map

exactly what my future will be. I guess that answers your question."

This is particularly interesting since it has been learned that the subject's husband is employed as an engineer in what seems to be a relatively substantial position.

SYMBOL IX. Aggression Symbol (Fig. 31).
Interpretation 1:

The fact that the stimulus-pattern was well structured was interpreted to mean that the subject does not express hostility explosively.

Q. "Are you considered to be a person with very much of a temper?"

A. "No I don't think that there is anything to be gained from outbursts."

The problem of ego-involvement might seem worthy of consideration in evaluating the last response. However, people who are quick to anger are often as proud of their spontaneity as others are of self-control.

Interpretation 2:

The characterization of the figure stated in Question 4 of the inquiry was interpreted as meaning that, when this subject was confronted by pressure to conform, she reacted quietly but with confidence in pursuing her own pattern.

Q. "How do you feel when situations seem to re-

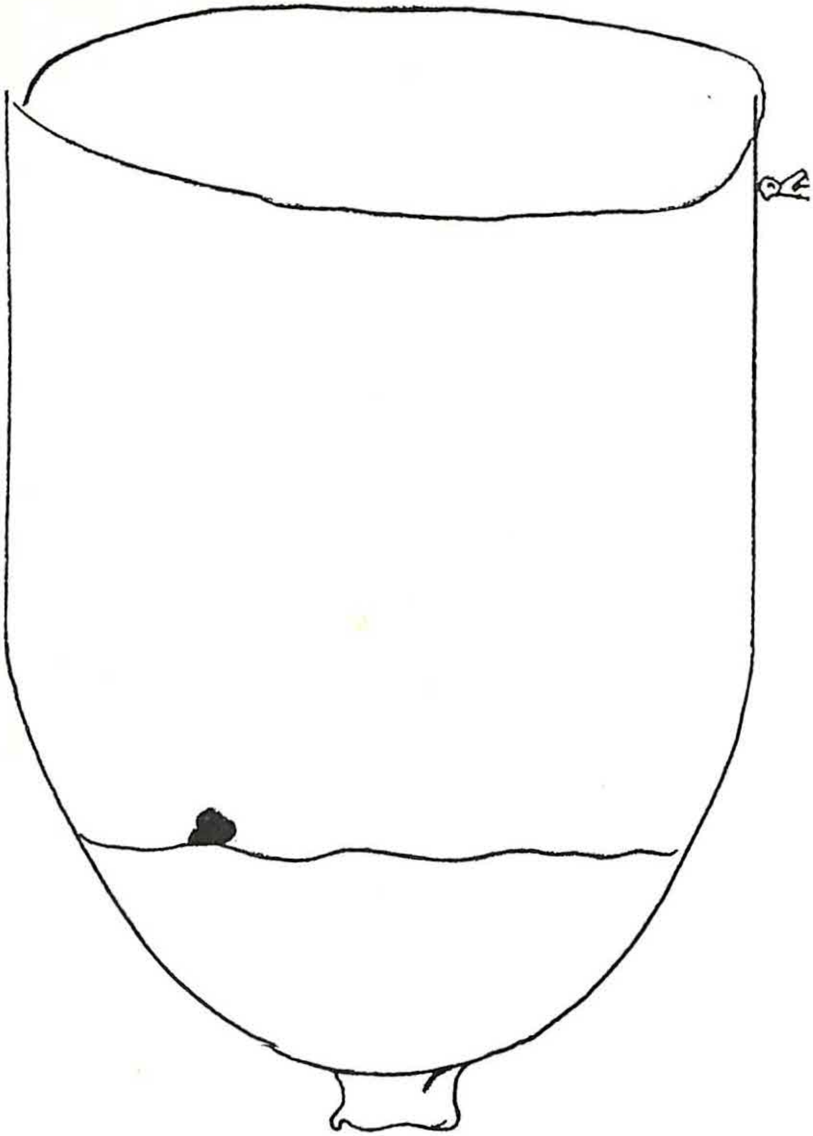


FIG. 32. Case B: Symbol X.

1. Flies around a brandy glass.
2. Easy.
3. Indifference.
4. None.
5. Hopelessness.
6. No.
7. Short-lived love affairs.

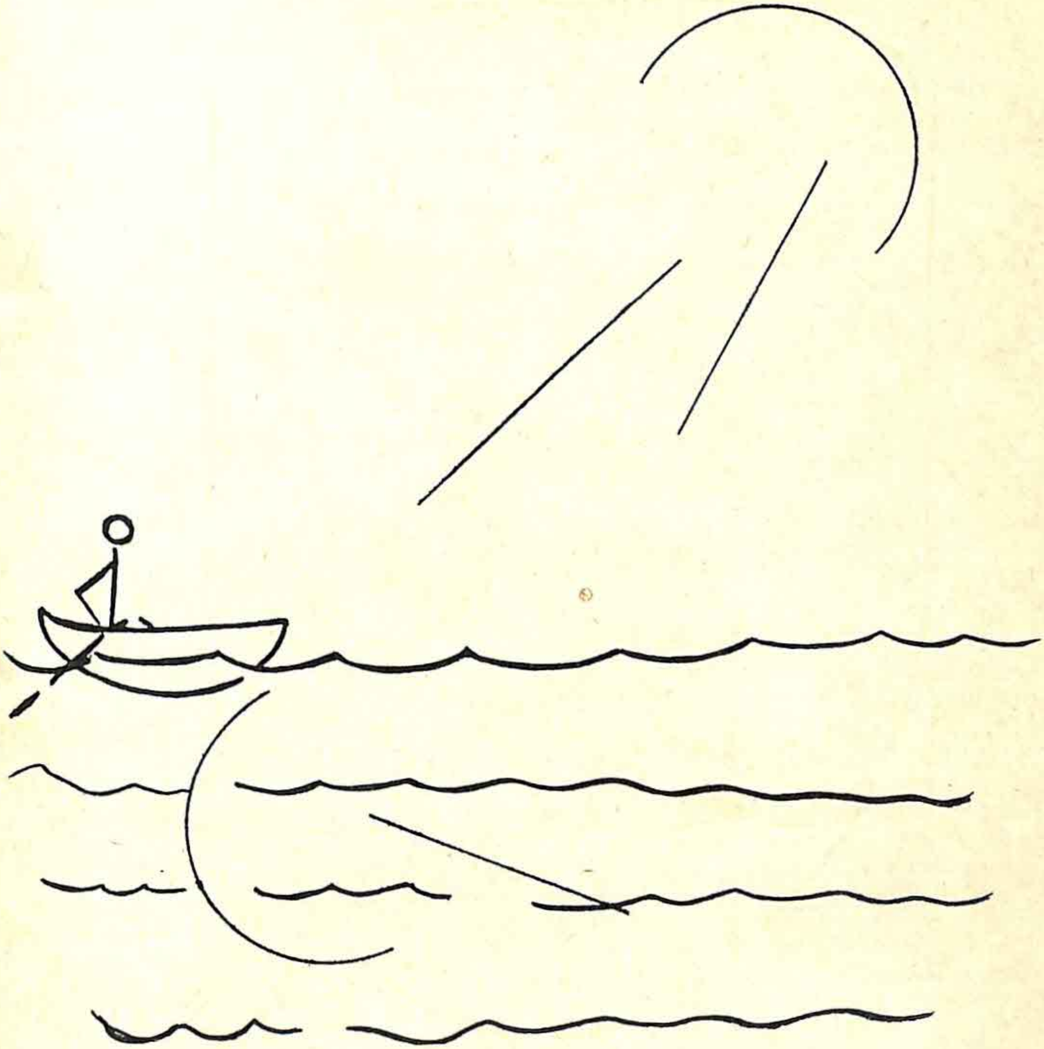


FIG. 33. Case B: Symbol XI.

1. Canoe on a moonlit lake.
2. Easy.
3. Weirdness.
4. He loves solitude, is not above rediscovering the magic usually captured on calendar pictures, will come home hungry and go to bed.
5. Pleasure.
6. Yes.
7. Cool summer nights.

quire that you change yourself to suit the desires of others?"

A. "You mean, was I one of the first to change to the 'new look'? Well, seriously, that was the thing that kept me from joining the WAC during the war. I am afraid I get pretty stubborn about liking things to be my own way. I don't care to tell others what to do, but one thing I just don't pay any attention to is their trying to dictate to me."

SYMBOL X. The Self-Concept (Fig. 32).

Interpretation 1:

The conception of a passive container holding something desirable which others seek, like flies near brandy, suggests that the subject sees herself as someone passive, and yet sought after. Passivity is further emphasized by the answer to Question 3 on the inquiry, "mood of indifference."

Q. "Generally speaking, do you consider yourself to be a 'go-getter' sort of person or not?"

A. "Uh *uh!* I'm a lazy-Susan. But I'm pretty lucky about things coming to me. Maybe that's why I don't really have to push myself too hard."

The last remark made further questioning unnecessary since it confirmed the entire interpretation.

SYMBOL XI. The Competitive Heterosexual Symbol (Fig. 33).

Interpretation 1:

From the fact that "solitude" appears only in connection with this symbol and Symbol 2 (individual heterosexual situation), it seemed that loneliness, rather than seclusion, was the connotation. (The latter was the interpretation on the basis of the inquiry in Symbol 2 alone.) Because the subject contrasts the mood of solitude with the attitude expressed outside of the home in Question 4 of the inquiry, a lack of emotional rapport with her husband seemed likely.

Q. "In intimate situations with your husband, have you sometimes felt as though he does not share your feelings?"

A. "Well, I'm not sure that I understand you. But most of the time I feel left out of things, as far as he is concerned. But I think that is just because we are so different."

Interpretation 2:

The substance of the answer to Question 4 on the inquiry seemed to mean that the subject had desired to experience extra-marital relations, but resisted; since the "magic" usually captured on calendar pictures seems to have a definite romantic and sexual connotation, and the subject states that "the appetite" is not satisfied.

Q. "Have you ever come close to finding out whether you could respond more completely to another man, I mean other than your husband?"

A. "Well, I haven't been unfaithful, if that's what you mean—although that's no promise for the future! (Laughs) Seriously though, I did plan one time, but I didn't go through with it."

The interpretation of this case illustrates the interpretation of the manner of drawing, the content, and the verbal associations of the subject. The verbal associations, particularly, helped in the successful analysis of this case. There was one "*P*+" rating. The rest of the interpretations were fully corroborated by the subject's remarks.

CHAPTER IV

RESULTS

A. ANALYSIS OF TOTAL RESULTS

THE results obtained in this study will be analyzed first in terms of the scoring categories. The object of this approach will be to learn whether the stimulus-patterns gave rise to interpretations which could be substantiated in a significant percentage of the cases. The second approach will aim to discover whether the stimulus-patterns were equally subject to validation. Attention will then be given to specific factors which may have a bearing on the results: the factors of sex, emotional stability, sequence of presentation, age, and cultural background.

The data, in terms of the scoring categories, are summarized in Table 3. All of the one hundred sixty-nine subjects are included. No more than three interpretations were made on a given subject-response. The average number was a total of twelve for the set of eleven stimulus-patterns. The 2,265 interpretations of all the subjects' responses served as

TABLE 3

INTERPRETATIONS FOR THE VARIOUS STIMULUS-PATTERNS IN TERMS OF THE RATING-CATEGORIES: ALL CASES

$N = 169$

Interpretations = 2,265

Symbol	Data	No Evidence "O"	Complete Corroboration "+"	Partial Corroboration "P"	Contradiction "–"	Partial Contradiction "P –"	Conflicting Evidence "±"	Total Number	Total Number Without "O"
Semi-circle	No.	13	203	12	18	1	2	249	236
	%	5.2	86.0	5.1	7.7	0.4	0.8		
	SE		2.4	1.4	1.8	0.4	0.6		
Semi-circle & single line	No.	19	160	4	14	0	4	201	182
	%	9.5	87.9	2.2	7.7		2.2		
	SE		2.4	1.1	1.9		1.1		
Straight line	No.	25	182	8	19	3	1	238	213
	%	10.5	85.4	3.8	8.9	1.4	0.5		
	SE		2.4	1.3	1.8	0.8	0.5		
Inverted "V's"	No.	38	152	7	9	1	0	207	169
	%	18.4	89.9	4.1	5.0	0.6			
	SE		2.3	1.5	1.7	1.8			
Diffuse mass	No.	22	152	3	4	1	1	183	161
	%	12.0	94.4	1.9	2.2	0.5	0.5		
	SE		2.9	0.91	1.1	0.5	0.5		
Crosslines	No.	34	146	9	7	4	2	202	168
	%	16.8	86.9	5.4	4.0	2.3	1.2		
	SE		3.1	2.2	1.8	1.5	1.0		
Counter-opposed semi-circles	No.	38	131	6	8	0	0	183	145
	%	20.8	90.3	4.1	5.5				
	SE		7.4	1.5	1.9				
Semi-circle within larger semi-circle	No.	19	144	4	12	1	4	184	165
	%	10.3	87.3	2.4	6.4	0.5	0.2		
	SE		3.2	0.3	1.8	0.5	0.3		
Jagged line	No.	23	153	10	8	1	2	197	174
	%	11.7	87.9	5.7	4.5	0.6	1.2		
	SE		2.3	1.8	1.6	0.6	0.8		
Incomplete "U"	No.	21	178	9	11	1	2	222	201
	%	9.4	88.6	4.4	5.5	0.5	1.1		
	SE		2.2	1.4	1.6	0.5	0.7		
Lines and semi-circles	No.	30	144	10	12	3	0	199	169
	%	15.1	85.7	5.9	7.1	1.7			
	SE		2.7	1.8	2.0	0.9			

the basis for the figures presented.

The percentage of interpretations for a given symbol which fell into each scoring category is shown. The first column contains the percentage of interpretative statements on which no evidence could be found in the validating material. The rest of the percentages were computed on the basis of the number of statements on which either supporting or negating evidence could be found. Thus it is possible to see the extent of agreement between the validating material and the interpretative statements for each symbol. From this table we can compare the relative frequency of the different scoring categories.

It is apparent on inspection of Table 3 that all the symbols offer confirmable interpretations in an overwhelming percentage of the cases. The percentages in the "plus" column range from 85.4% to 94.7%. The size of the percentages assumes further importance in the fact that the standard errors of those in the "plus" column are all less than one-thirteenth of the proportion, and in most cases equal about one thirty-fifth of the proportion. This makes the probability of finding complete agreement for interpretations based upon the subjects' responses very high.

Considering the data in terms of negative probability (see "minus" column), it appears from Table 3 that interpretations based on the straight-line (Fig. 4) pattern were most often contradicted by the validating material. Here the contradicted statements amount to 8.9%; the partially contradicted statements amount to 1.4%; and those for which there was both supporting and contradictory evidence amount to 0.5%. If we multiply the sigmas of each proportion by three and add these numbers to the original percentages, we have for each case the maximum percentage likely to occur in the negative results (counting the "plus-minus" group on a half basis). This computation would indicate a probable maximum of 18.98%. Thus the most extreme case of lack of support still leaves a high degree of confidence in the reliability and validity of the procedure involved in this test.

TABLE 3A

INTERPRETATIVE STATEMENTS FOR EACH SYMBOL FALLING INTO THE VARIOUS SCORING GROUPS: ALL CASES

Scoring Categories Used for Validation	Per cent of Cases in Each Scoring Category
0	0– 44
+	50–100
P+	0– 53
±	0– 9
P±	0– 0.3
–	0– 30.7
P–	0– 11

The effect of computing percentages without counting the cases in which no evidence was available is of interest to us. A glance at Table 3 indicates that the highest figure in the "No Evidence" column was 20.8%. Since this means that, even here, judgments could be rendered in 145 cases, it is evident that an adequate sample was employed.

The ranges in Table 3A initiated to show the spread of scoring categories for individual cases, invite caution. There was only one case in which 30.7% of the judgments were contradicted, the remaining statements having been wholly corroborated. It is noteworthy too that the lowest figure of plus-ratings in any one case was 50%. There was partial corroboration in 32% of the plus-rating

TABLE 4

INTERPRETATIONS IN THE VARIOUS RATING-CATEGORIES IN TERMS OF THE STIMULUS-PATTERNS: ALL CASES

Symbol	Data	No Evidence "O"	Complete Corroboration "+"	Partial Corroboration "P+"	Contradiction "−"	Partial Contradiction "P−"	Conflicting Evidence "±"	Total Number	Total Number Without "O"
Semi-circle	No.	13	203	12	18	1	2	249	236
	%	4.6	11.6	14.6	14.8	6.2	11.2	11.0	11.9
	SE		0.8	3.9	3.2	6.0	7.0		
Semi-circle & single line	No.	19	160	4	14	0	4	201	182
	%	6.7	9.2	4.9	11.5		22.2	8.9	9.2
	SE		0.7	2.2	2.9		9.7		
Straight line	No.	25	182	8	19	3	1	238	213
	%	8.8	10.4	9.8	15.6	18.8	5.6	10.5	10.7
	SE		0.7	3.3	3.3	9.8	5.6		
Inverted "Vs"	No.	38	152	7	9	1	0	207	169
	%	13.5	8.7	8.5	7.4	6.2		9.1	8.5
	SE		0.7	3.0	2.3	6.0			
Diffuse mass	No.	22	152	3	4	1	1	183	161
	%	7.8	8.7	3.7	3.3	6.2	5.6	8.1	8.1
	SE		1.0	2.1	1.5	6.0	5.6		
Crosslines	No.	34	146	9	7	4	2	202	168
	%	12.0	8.4	10.9	5.7	25.0	11.2	9.0	8.5
	SE		0.8	3.5	2.1	11.0	7.0		
Counter-opposed semi-circles	No.	38	131	6	8	0	0	183	145
	%	13.4	7.5	7.3	6.6			8.1	7.3
	SE		1.0	2.9	2.3				
Semi-circle within larger semi-circle	No.	19	144	4	12	1	4	184	165
	%	6.7	8.2	4.9	9.8	6.2	22.2	8.1	8.3
	SE		0.3	2.2	2.7	6.0	9.7		
Jagged line	No.	23	153	10	8	1	2	197	174
	%	8.2	8.8	12.2	6.6	6.2	11.2	8.7	8.8
	SE		0.7	3.4	2.3	6.0	7.0		
Incomplete "U"	No.	21	178	9	11	1	2	222	201
	%	7.4	10.2	10.9	9.0	6.2	11.2	9.8	10.1
	SE		0.7	3.3	2.6	6.0	7.0		
Lines and semi-circles	No.	30	144	10	12	3	0	199	169
	%	10.6	8.2	12.2	9.8	18.8		8.8	8.5
	SE		0.6	3.4	2.7	9.8			
	Total No.	282	1745	82	122	16	18	2265	1983

judgments. The highest percentage of no evidence ratings occurred in the Indian group, and in one case it was as high as 44%. The special conditions obtaining here have been described elsewhere. However, more than half the judgments could still be evaluated, even in the case of this individual.

Table 4 deals with other problems which the total data help us solve. Here the percentages have been computed so as to compare each symbol with each of the others in terms of the contribution it makes to the total number of each category. By comparing the percentage in each column with the percentages in the last (or total) column we can see the relationship between this contribution and the percentage of interpretations indicated for each symbol.

The most striking fact which this table offers is that all the symbols are equally valid. The percentages in the "plus-column" are practically identical with those in the last. In other words, while the symbols vary slightly in the number of interpretations to which they readily lend themselves in an individual record, there is practically no difference between their likelihood to produce a given number of interpretations and the likelihood of these interpretations' being wholly substantiated. Variations exist in the other categories, but they do not affect this relationship.

Some of these variations are interesting to examine. For example, the "femaleness" pattern (Fig. 1) was the easiest pattern to relate to the validating material (i.e., had the lowest per cent in the *"O"* or "no evidence" column). This is probably a corollary of the degree of attention given the "femaleness" pattern (Fig. 1) by most of the intensive studies of individual behavior. The percentage of "no evidence" ratings for the "maleness" pattern (Fig. 4), was practically double that for the "femaleness" pattern (Fig. 1).

The most difficult symbols to validate in our study were those purporting to deal with "maleness" (Fig. 4), "femaleness" (Fig. 1), competitive heterosexuality (Fig. 7), and sexual anxiety (Fig. 10). The highest figure in the "no evidence" column on any symbol was 13.5%, and the lowest number of judgments on any symbol turned out to be 145. This is, of course, large enough to provide a reliable basis for conclusions.

The reason that these patterns appear to refer to more obscure material than do the others could be found in the fact that with these patterns we enter the field of tabooed social values. The symbolism involved in the remaining patterns could hence be more easily validated than could these four patterns. Nevertheless, it will be observed that substantial confirmation of validity has been attained even for these patterns as the + and *P*+ figures in Table 5 clearly show. Incidentally, the remaining symbols contributed to the "no evidence" column in proportion to the number of judgments based upon them in the entire study.

Consideration must be given to the "minus" column at this point. The only three symbols for which we find disproportionately higher percentages in this column than in the total column are the symbols for intimate heterosexuality (11.5%), "femaleness" (14.8%), and "maleness" (15.6%). In view of the fact that these three symbols have approximately the same percentages in the "plus" column as in the total column, it would seem that the author might have felt encouraged to extend her conclusions beyond the evidence available. The fact that these are respectively fourth, first, and second highest in number of interpretations derived from the completed patterns, however, tends to support the conclusion that, in spite of the figures in the "minus" column, these proved to be exceedingly productive and reliable patterns.

In considering the conflictual scoring of "plus-minus," it is worth noting that the concentric circles pattern (Fig. 3) and the intimate heterosexual symbol (Fig. 6) comprise each 22% of the items thus categorized. The conflictual rating may have been due in these cases to confusing statements, leading to uncertain interpretations, by independent workers; or it may be that it is particularly easy in these two cases to emphasize an aspect of behavior brought out by the drawings and inquiry, without attributing sufficient importance to coexisting factors. Similar explanations are probably appropriate with reference to the less markedly skewed figures for the aggression (Fig. 11) and the "maleness" (Fig. 4) symbols. The other figures in this colmn are not significant when compared to the base-percentages in the last column of Table 4. In most cases, there were no ratings in this category at all, showing that obtainable

TABLE 5

INTERPRETATIVE STATEMENTS ON EACH SYMBOL FALLING INTO THE VARIOUS SCORING GROUPS: TOTAL NUMBER, MALES

Symbol	Data	No Evidence "O"	Complete Corroboration "+"	Partial Corroboration "P +"	Conflicting Evidence "±"	Contradiction "−"	Partial Contradiction "P −"	Total Number	Total Number Without "O"
Semi-circle	No.	12	80	10	4	6		112	100
	%	10.0	80.0	10.0	4.0	6.0			
	SE		4.0	3.0	2.0	2.4			
Semi-circle & Single line	No.	12	83	8	1	6		110	98
	%	10.9	84.7	8.2	1.2	6.1			
	SE		3.0	2.8	1.0	2.4			
Straight line	No.	13	87	14	5	14	1	134	121
	%	9.7	71.9	11.6	4.1	11.6	8.0		
	SE		4.0	9.0	1.8	9.0	0.9		
Inverted "Vs"	No.	20	72	11		7	1	111	91
	%	18.0	79.1	12.1		7.7	1.1		
	SE		4.0	3.0		2.8	1.0		
Diffuse mass	No.	26	80	5	1	4		116	90
	%	16.2	89.0	5.6	1.1	4.5			
	SE		3.0	2.0	1.0	2.0			
Crosslines	No.	21	97	1		5		124	103
	%	16.9	94.2	0.9		4.9			
	SE		2.0	9.7		2.0			
Counter-opposed semi-circles	No.	16	74	15		4		109	93
	%	17.0	79.6	16.1		4.3			
	SE		4.0	3.0		2.0			
Semi-circle within larger semi-circle	No.	8	82	3	1	4	1	99	91
	%	8.0	90.1	3.3	1.1	4.4	1.1		
	SE		3.0	1.7	1.0	2.0	1.0		
Jagged line	No.	13	87	3		6		106	93
	%	12.2	93.5	3.3		6.5			
	SE		2.4	1.7		2.4			
Incomplete "U"	No.	18	106	2	1	5		132	114
	%	13.5	93.0	1.8	0.9	4.4			
	SE		2.0	1.4	0.9	1.8			
Lines and semi-circles	No.	17	94	5		5	3	124	107
	%	13.7	87.8	4.7		4.6	2.8		
	SE		3.0	2.1		2.1	1.7		

evidence tended to be clear on the points raised in the experimental interpretations, even though this evidence was sometimes not conclusive as in *P*+ and *P*− ratings.

The data dealing with conflicting evidence and partial contradiction, involving as they do together 34 of the 2,265 interpretations, are negligibly small and do not require detailed analysis.

B. SEX FACTORS

Tables 5 and 6 summarize the results according to the percentage of interpretative statements for each symbol in each scoring group. Table 5 contains these results for males only. Table 6 contains the results for females only.

In Table 5, the only figure in the "plus" column that is below 79% is that for the "maleness" pattern (Fig. 4). Since this is one of the highest of the corresponding figures in Table 6, while that for the "femaleness" pattern (Fig. 1) is lower in the female group, it seems likely that the difficulty of separating the personality's identification with the parental role from his conception of that pattern causes some difficulty in analysis. The fact that many of these cases were analyzed without benefit of inquiry must be kept in mind, however, in interpreting this point. In spite of this, the "maleness" pattern (Fig. 4) in the "plus" column of Table 5 is represented by 71.9% with a sigma of ±4%. In the "femaleness" pattern (Fig. 1) row of Table 6, the datum in the "plus" column is 81% with a sigma of ±4%. These, then, are highly reliable results. In general, there seems to be greater variability in the correctness of interpretations in the male group than in the female group, where the percentages are neither as high nor as low as the extremes of the male group. Why this should be so is not clear. However, the effectiveness of the data for both sexes is so apparent that this slight variation should not matter.

The fact that the percentages for the two sexes go down when relationships with the opposite sex as a group are considered has been discussed above. It is clear, in these tables, that evidence is less easy to find in these areas, although the relationships with those of the same sex-group are rather easily determined from biographical data. Thus, in both sexes, the percentage in the "plus" column go up for the same sex group, and those in the "no-evidence" and "*P*-plus" column rise when the opposite sex is considered as a group. Again, it is important to note that nowhere do the figures in the "plus" column become unsubstantial or insignificant.

The only two relatively large figures in the "minus" column in both tables

TABLE 6

Interpretative Statements on Each Symbol Falling into the Various Scoring Groups: Total Number, Females

Symbol	Data	No Evidence "O"	Complete Corroboration "+"	Partial Corroboration "+P"	Conflicting Evidence "±"	Partial Conflicting "P±"	Contradiction "—"	Partial Contradiction "P—"	Total Number	Total Number Without "O"
Semi-circle	No. % SE	3 2.9	81 81.0 4.1	5 5.0 2.0			14 14.0 3.5		103	100
Semi-circle & single line	No. % SE	15 15.1	73 89.0 3.5	1 1.0 1.0	1 1.0 1.0		6 7.3 9.0	1 1.0 1.0	97	82
Straight line	No. % SE	12 11.2	85 89.5 3.0	2 2.1 1.4	1 1.0 1.0		5 5.2 2.0	2 2.1 1.4	107	95
Inverted Vs"	No. % SE	18 14.4	97 90.6 2.6	4 3.7 1.7			6 5.6 2.0		115	107
Diffuse mass	No. % SE	9 9.4	70 81.4 4.1	8 9.3 3.4			7 8.1 3.0	1 1.2 1.0	95	86
Crosslines	No. % SE	22 23.9	61 87.1 5.2	6 6.6 9.3	1 1.1 1.3		1 1.1 1.3	1 1.1 1.3	92	70
Counter-opposed semi-circles	No. % SE	17 11.0	70 85.4 3.7	5 5.8 2.6		1 1.1 1.1	6 6.9 8.7		99	82
Semi-circle within larger semi-circle	No. % SE	8 8.6	72 84.7 3.9	5 5.8 2.4	2 2.3 5.0		6 7.0 2.8		93	85
Jagged line	No. % SE	12 13.3	62 81.6 4.3	8 10.2 3.4	2 2.5 5.6		4 5.1 2.6		88	76
Incomplete "U"	No. % SE	13 13.0	74 84.1 3.7	7 8.0 2.8			6 6.8 8.4	1 1.1 1.0	101	88
Lines and semi-circles	No. % SE	16 17.7	64 86.5 4.2	5 6.5 2.8			5 6.5 2.8		90	74

TABLE 7

INTERPRETATIVE STATEMENTS FOR EACH SYMBOL FALLING INTO THE VARIOUS SCORING GROUPS: TOTAL NUMBER OF ADULT PSYCHIATRIC CASES

Symbol	Data	No Evidence "O"	Complete Corroboration "+"	Contradiction "P−"
Semi-circle	No. % SE	4 4.5	79 94.0 3.6	5 6.0 2.7
Semi-circle & single line	No. % SE	9 12.5	58 92.0 3.4	5 7.0 3.4
Straight line	No. % SE	6 7.1	70 89.0 3.6	8 1.0 3.6
Inverted "Vs"	No. % SE	6 9.3	64 94.0 2.9	4 6.0 2.9
Diffuse mass	No. % SE	8 11.4	55 89.0 3.9	7 11.0 3.9
Crosslines	No. % SE	8 10.2	67 95.0 2.6	3 4.3 2.6
Counter-opposed semi-circles	No. % SE	9 5.8	61 94.0 2.9	4 6.1 2.9
Semi-circle within larger semi-circle	No. % SE	4 6.7	50 91.0 3.8	5 9.0 3.8
Jagged line	No. % SE	7 10.9	58 96.6 2.5	2 3.3 2.5
Incomplete "U"	No. % SE	3 3.3	83 95.0 2.3	4 4.7 2.4
Lines and semi-circles	No. % SE	10 13.5	61 95.0 2.8	3 4.7 2.8

occur in connection with the symbol for the parent of the same sex (Fig. 1 or Fig. 4). This is interesting because it gives support to the previously noted trend in the discussion the "plus" column, namely, that the confusion of the actual parental figure with the role assumed by the individual who identifies with that parent tends to lower the accuracy of interpretation.

C. FACTOR OF EMOTIONAL STABILITY

Those who doubt the existence of generalized (generic) symbols have occasionally pointed out that the pioneer work with symbols of this sort was done by researchers interested in pathology. For this reason, we include here Tables 7, 8, 9, and 10. Table 7 refers to 60 adult psychiatric cases, while Table 8 refers to 50 cases of apparently normal adults.

TABLE 8

INTERPRETATIVE STATEMENTS FOR EACH SYMBOL FALLING INTO THE VARIOUS SCORING GROUPS: TOTAL NUMBER, NORMAL ADULTS

Symbol	Data	No Evidence "O"	Complete Corroboration "+"	Partial Corroboration "P+"	Contradiction "−"	Total Number	Total Number Without "O"
Semi-circle	No. % SE	4 5.2	66 92.0 3.0		6 8.3 3.0	76	72
Semi-circle & single line	No. % SE	6 14.6	44 97.2 2.8		1 2.8 2.8	51	45
Straight line	No. % SE	5 8.8	48 92.3 3.4	2 3.8 2.4	2 3.8 2.4	57	52
Inverted "Vs"	No. % SE	13 25.0	36 92.3 4.2		3 7.7 4.2	52	39
Diffuse mass	No. % SE	9 17.0	43 97.7 2.0		1 2.3 2.0	53	44
Crosslines	No. % SE	8 18.3	44 100.0			52	44
Counter-opposed semi-circles	No. % SE	7 14.6	41 100.0			48	41
Semi-circle within larger semi-circle	No. % SE	3 5.4	53 100.0			56	53
Jagged line	No. % SE	1 2.1	46 96.0 2.8		2 4.0 2.8	49	48
Incomplete "U"	No. % SE	1 1.8	51 96.5 2.6		2 3.5 2.6	54	53
Lines and semi-circles	No. % SE	2 4.4	42 97.5 2.2		1 2.5 2.2	45	43

The only difference between these two tables is that the "no evidence" column in the case of the normals is higher in some categories, and so are the percentages of "pluses."

Tables 9 and 10 summarize the same data pertaining to children. Table 9, referring to children with behavior disorders, is not significantly different from Table 10 (normal children). The discrepancies here are easily explained by the fact that normal group drawings were all analyzed without access to inquiry.

In other words, the degree of reliability was slightly higher in the normal group. This may be due to the fact that the attitudes of those in need of, or

TABLE 9

INTERPRETATIVE STATEMENTS FOR EACH SYMBOL FALLING INTO THE VARIOUS SCORING GROUPS: TOTAL NUMBER, BEHAVIOR PROBLEM CHILDREN

Symbol	Data	No Evidence "O"	Complete Corroboration "+"	Partial Corroboration "P+"	Conflicting Evidence "±"	Contradiction "—"
Semi-circle	No.		23	1		1
	%		92.0	4.0		4.0
	SE		5.0	4.0		4.0
Semi-circle & single line	No.	3	17			1
	%	14.0	95.0			5.0
	SE		5.0			5.0
Straight line	No.		20		2	3
	%		80.0		8.0	12.0
	SE		4.5		0.5	3.5
Inverted "Vs"	No.	1	22			2
	%	4.0	91.0			9.0
	SE		6.0			6.0
Diffuse mass	No.	3	20			1
	%	13.0	95.0			5.0
	SE		5.0			5.0
Crosslines	No.	4	21			
	%	16.0	100.0			
	SE					
Counter-opposed semi-circles	No.	3	16			
	%	16.0	100.0			
	SE					
Semi-circle within larger semi-circle	No.	1	21			
	%	5.0	100.0			
	SE					
Jagged line	No.	2	20			
	%	9.0	100.0			
	SE					
Incomplete "U"	No.	4	18			1
	%	17.0	95.0			5.0
	SE		5.0			5.0
Lines and semi-circles	No.	3	20			1
	%	12.0	95.0			5.0
	SE		5.0			5.0

undergoing, therapy are in a greater state of flux than are the attitudes of those who are relatively stable. The reason for the larger number of *"O"* scorings in the normal group is probably that half of the cases were from the interview group in which evidence was hardest to establish on some points. For example, sexual anxiety (Fig. 10) was sometimes difficult to interpret in terms of concrete behavior, and thus the necessary information (positive or negative) was comparatively difficult to elicit. Substantially, these four tables do establish the fact that the symbols are about equally appropriate for the emotionally disturbed and the more adequately functioning members of a given population.

TABLE 10

INTERPRETATIVE STATEMENTS FOR EACH SYMBOL FALLING INTO THE VARIOUS SCORING GROUPS: TOTAL NUMBER, NORMAL CHILDREN

Symbol	Data	No Evidence "O"	Complete Corroboration "+"	Partial Corroboration "P+"	Conflicting Evidence "±"	Contradiction "—"
Semi-circle	No. % SE	3 19.4	38 88.4 4.7	(13.9% P+)	1 2.3 7.3	4 9.3 (2.3% P−) 4.3
Semi-circle & single line	No. % SE	8 19.4	29 85.4 9.0	(9.0% P+)	3 8.8 4.7	2 5.9 1.0
Straight line	No. % SE	8 16.6	36 90.0 4.5	(15.0% P+)		4 10.0 (2.5% P−) 4.5
Inverted "Vs"	No. % SE	12 34.2	21 90.5 1.0	(18.1% P+)		1 4.5 1.0
Diffuse mass	No. % SE	2 6.6	28 90.3 5.2	(10.0% P+)	1 3.2 0.3	2 6.5 (3.0% P−) 1.0
Crosslines	No. % SE	9 27.3	20 83.3 5.5	(16.6% P+)	1 4.2 4.0	3 12.5 6.7
Counter-opposed semi-circles	No. % SE	15 35.5	24 88.9 6.0	(15.5% P+)		3 11.1 6.0
Semi-circle within larger semi-circle	No. % SE	5 12.6	29 80.6 6.5	(10.0% P+)	2 5.5 3.7	5 13.9 (2.5% P−) 5.2
Jagged line	No. % SE	3 8.8	27 87.1 6.0	(22.0% P+)	1 3.2 3.0	3 9.7 5.2
Incomplete "U"	No. % SE	10 20.4	31 81.6 6.2	(15.0% P+)	1 2.6 2.4	6 15.9 5.9
Lines and semi-circles	No. % SE	8 21.6	28 96.6 3.3	(28.0% P+)		1 3.4 (3.0% P−) 3.3

D. FACTOR OF AGE

Table 11 contains the results computed in percentage for the group of 46 children, age range 6-16. Table 12 contains the results computed similarly for the adult group, age range 17-69. When these two tables are compared, the differences appear very slight. It will be observed that percentages in the children's group have been influenced by the comparative difficulty of obtaining judgments without direct inquiry on most cases.

Table 13 summarizes the results for the 12 children in the Indian groups. The size of the sigmas indicates a relative decrease of validity. However, there is some interesting information to be gleaned from this table. In the first place, *the*

TABLE 11

INTERPRETATIVE STATEMENTS FOR EACH SYMBOL FALLING INTO THE VARIOUS SCORING GROUPS: TOTAL NUMBER, CHILDREN

Symbol	Data	No Evidence "O"	Complete Corroboration "+"	Partial Corroboration "P+"	Conflicting Evidence "±"	Contradiction "−"	Partial Contradiction "P−"	Total Number	Total Number Without "O"
Semi-circle	No. % SE	3 4.2	55 80.1 4.8	7 10.3 3.6	1 1.4 1.4	4 5.9 2.9	1 1.4 1.4	71	68
Semi-circle & single line	No. % SE	11 17.5	43 82.7 5.2	3 5.7 3.2	3 5.7 3.2	3 5.7 3.2		63	52
Straight line	No. % SE	8 10.9	50 77.0 5.2	6 9.2 3.6	2 3.1 2.2	6 9.2 3.6	1 1.5 1.5	73	65
Inverted "Vs"	No. % SE	13 22.0	39 84.7 5.3	4 8.7 4.2		3 6.5 3.7		59	46
Diffuse mass	No. % SE	5 8.8	45 79.0 5.6	3 5.3 3.1	1 1.7 2.0	2 3.5 2.6	1 1.7 2.0	57	52
Crosslines	No. % SE	13 22.0	27 77.1 7.1	4 11.4 5.3	1 2.9 3.0	3 8.6 4.6		48	35
Counter-opposed semi-circles	No. % SE	18 29.5	35 81.5 5.9	5 11.6 4.7		3 7.0 3.8		61	43
Semi-circle within larger semi-circle	No. % SE	6 9.4	47 81.0 5.0	4 6.9 0.6	2 3.4 0.6	4 6.9 0.6	1 1.7 0.5	64	58
Jagged line	No. % SE	5 8.9	40 78.5 5.6	7 13.7 4.7	1 1.9 0.6	3 5.9 3.4		56	51
Incomplete "U"	No. % SE	14 19.7	43 75.5 5.1	6 10.6 3.7	1 1.7 2.9	7 12.3 3.3		71	57
Lines and semi-circles	No. % SE	11 10.0	89 90.5 9.4	8 8.1 4.3		1 1.0 0.3	1 1.0 0.3	110	99

children involved were five years younger at the time the validating material had been collected than they were at the time the test material was administered. This represents a difference between the ages of eight and thirteen. The greatest lowering of percentages in the "plus" column occurred with reference to the following symbols:

(1) Individual heterosexual situation
(2) Peer groups
(3) Regression and mother tie
(4) Self-concept
(5) Competitive heterosexual situations

It seems evident that these are the areas in which one would logically expect much change over this period of five years.

The reason for the relatively high predictive value (85% "plus") of the sexual anxiety symbol may be the fact that the manner of handling, rather than a specific source of anxiety, was emphasized in the analyses. This might not change very much with age. Of course, the other data do not shed light on this problem, since all that can be told from them is that the symbol is predictive at the various age levels also for members of our own cultural group.

Although the sigmas are relatively large (due to the relatively small number of cases), in comparison with the data from the rest of this study (in which one

TABLE 12

INTERPRETATIVE STATEMENTS FOR EACH SYMBOL FALLING INTO THE VARIOUS SCORING GROUPS: ALL ADULTS

Symbol	Data	No Evidence "O'	Complete Corroboration "+"	Partial Evidence (Corroboratory) "P+"	Contradiction "—"	Total Number	Total Number Without "O"
Semi-circle	No.	8	145		11	164	156
	%	12.5	93.0		7.0		
	SE		2.0		2.0		
Semi-circle & single line	No.	15	102		6	123	118
	%	12.2	87.0		13.0		
	SE		3.0				
Straight line	No.	11	118	2	10	141	130
	%	7.9	90.1	1.5	7.7		
	SE		2.2	1.4	2.0		
Inverted "Vs"	No.	19	100		7	126	107
	%	15.0	93.3		6.7		
	SE		2.0				
Diffuse mass	No.	17	98		8	123	106
	%	13.8	92.5		7.5		
	SE		2.1				
Crosslines	No.	16	111		3	130	114
	%	12.4	97.4		2.6		
	SE		1.5				
Counter-opposed semi-circles	No.	16	102		6	124	108
	%	12.9	95.3		4.7		
	SE		1.7				
Semi-circle within larger semi-circle	No.	7	103		5	115	108
	%	6.1	96.2		3.8		
	SE		1.7				
Jagged line	No.	8	104		4	116	108
	%	6.9	96.3		3.7		
	SE		1.7				
Incomplete "U"	No.	4	134		6	144	140
	%	2.8	96.7		3.3		
	SE		1.6				
Lines and semi-circles	No.	12	103		4	119	107
	%	10.1	96.4		3.6		
	SE		1.6				

becomes accustomed to sigmas of one-thirtieth of the proportion), those for the stable areas are of good value in the Indian Group. They are about one-eighth of the proportion, and thus indicate a better-than-chance distribution.

Results on the twelve Indian children are probably inconclusive. On the other hand, since they indicate high reliability, they may be regarded as contributory evidence, from the viewpoint of the test as a whole. On the other hand, when these results are coupled with the findings of anthropological studies such as Roheim's and Hallowell's, they may not necessarily prove the universality of the symbols selected, but the likelihood of their doing so is greatly enhanced.

E. Stimulus Sequence as a Factor

Approximately one-third of the subjects were given the symbols in random order. These results are summarized in Table 14. Comparing them with the results in Table 3, the "plus" percentages seem to be slightly higher. The explanation for this might easily lie in the elimination of thirty cases in which no inquiry had been conducted. The order was the same for everyone in this particular group (normal children). In any case, the results are not impressively higher in Table 11. The conclusion one can draw is that, even if the only reason for less than 100% plus-ratings for all symbols were a question of contamination of attitudes, the 15% or less devia-

TABLE 13

Interpretative Statements for Each Symbol Falling into the Various Scoring Groups: Indian Group

Symbol	Data	No Evidence "O"	Complete Corroboration "+"	Conflicting Evidence "±"	Contradiction "−"
Semi-circle	No. % SE	2 15.4	9 81.8 (54.5% P+) 11.3		2 18.1 11.3
Semi-circle & single line	No. % SE	0 0	8 66.6 (8.3% P+) 13.5		4 33.3 13.5
Straight line	No. % SE	4 21.1	12 80.0 10.0		3 20.0 (13.3% P−) 10.0
Inverted "Vs"	No. % SE	6 30.0	11 85.7 (21.4% P+) 10.0		2 14.2 (7.1% P−) 10.0
Diffuse mass	No. % SE	3 21.4	10 90.9 (45.5% P+) 8.4		1 9.0 8.4
Crosslines	No. % SE	7 48.9	7 70.0 (45.5% P+) 15.0		3 30.0 (27.3% P−) 15.0
Counter-opposed semi-circles	No. % SE	5 37.5	5 62.5 (40.0% P+) 14.8	1 12.5 (20.0% P) 10.0	2 25.0
Semi-circle within larger semi-circle	No. % SE	3 21.4	7 63.6 (27.3% P+) 14.5	1 9.0 8.5	3 27.2 13.5
Jagged line	No. % SE	2 20.0	8 88.8 (37.5% P+) 10.5		1 11.1 (12.5% P−) 10.5
Incomplete "U"	No. % SE	5 27.8	9 64.2 (23.1% P+) 12.7	3 21.4 (15.5% P) 11.0	2 14.2 (7.7% P−) 9.0
Lines and semi-circles	No. % SE	2 10.5	11 64.7 (11.8% P+) 11.6		6 35.3 (11.8% P−) 11.6

TABLE 14

INTERPRETATIVE STATEMENTS FOR EACH SYMBOL FALLING INTO THE VARIOUS SCORING GROUPS: RANDOM ORDER GROUP

Symbol	Data	No Evidence "O"	Complete Corroboration "+"	Partial Corroboration "P"	Conflicting Evidence "±"	Contradiction "−"	Total Number	Total Number Without "O"
Semi-circle	No. % SE	3 4.1	62 88.6 3.8	1 1.4 1.2		7 10.0 3.6	73	70
Semi-circle & single line	No. % SE	5 8.9	49 94.3 3.2			3 5.7 3.2	57	52
Straight line	No. % SE	5 8.9	52 91.2 3.8		1 1.8 3.4	4 7.7 3.4	62	57
Inverted "Vs"	No. % SE	8 12.9	50 92.5 3.6	1 1.9 1.5		3 5.8 3.2	62	54
Diffuse Mass	No. % SE	4 7.7	48 90.8 3.9	1 1.5 1.6		4 7.5 3.4	57	53
Crosslines	No. % SE	8 1.3	51 98.1 1.9			1 1.9 1.9	60	52
Counter-opposed semi-circles	No. % SE	7 11.8	49 94.2 3.2	1 1.9 1.9		1 3.9 6.7	59	52
Semi-circle within larger semi-circle	No. % SE	4 7.5	46 93.8 3.3			3 6.2 3.3	53	49
Jagged line	No. % SE	2 3.8	47 94.0 3.3			3 6.0 3.3	52	50
Incomplete "U"	No. % SE	1 1.5	63 100				64	63
Lines and semi-circles	No. % SE	3 5.0	52 92.8 3.4			4 7.2 3.4	59	56

tion from 100% may be attributed to the significance of the order in which the stimulus-patterns were presented. However, there is good reason to believe, as has already been demonstrated, that such factors as lack of sufficient inquiry and lack of inclusiveness of validating data affecting Table 3, all have a negative effect on the "plus" percentages. Thus one is left with the impression that the symbols included in this investigation are sufficiently unique in individual meaning, and effective as individual stimuli, to reduce the sequence of their presentation to a position of relative unimportance. We can thus conclude that the order of presentation of stimuli—no less than the factors of sex, emotional stability, age, and cultural background—has no appreciable effect on the results obtained by our technique.

CHAPTER V

CONCLUSIONS

A. GENERAL IMPLICATIONS OF THIS STUDY

THE hypothesis tested in this study was that certain abstract visual forms may be regarded as universal symbols. Eleven forms were selected for use as stimulus-patterns. The reactions of the subjects to these patterns, expressed on both a non-verbal and a verbal level, were analyzed in terms of the hypothetical meanings of the stimulus-patterns. Validation of the meanings of the stimulus-patterns depended upon whether these interpretations about the subjects could be established as true.

The results described in detail in the preceding chapter show that the eleven stimulus-patterns all disclosed high predictability as cues to emotional dynamics. We may thus conclude that these stimulus-patterns could be reliably interpreted on the assumption that they contain inherently those subtle symbolic meanings which we attributed to them in our basic, pre-experimental rationale.

The results are similar for both emotionally stable and emotionally unstable members of our culture, regardless of sex or age. They are less conclusive, because of the small sample used, for members of Indian cultures. In the Indian groups, too, however, the evidence supports the hypothesis under question. Further work with members of other cultures is therefore warranted.

B. USE OF THE SYMBOLS AS BASIS FOR PROJECTIVE TECHNIQUE

In discussing the use to which the symbols may be put in a projective test, it is necessary first to consider the meaning of projection in this sense. As understood by this writer, projection refers to the qualities which a situation assumes as a result of the dynamics which the subject brings to bear on it. The more plastic a situation is, the more unstructured it is, the more of his own qualities may we expect a subject to reveal in coping with the situation. However, a subject's "projections" may be observed in all life situations. In using a *projective technique,* we standardize a set of conditions which can be presented to each new subject, so that his reactions in this laboratory situation may be compared to the reactions of others. Even in so thoroughly structured a situation as one in which a soldier is required to give a salute, we may see this. Some men respond snappily and with pride in the act. Others respond so quickly that the salute is barely meaningful. Still others are slow and deliberate. In each case the individual shows his own style of reaction and, thus, certain aspects of his personality.

The Rorschach test is possibly the least structured and the most plastic of the various projective techniques in common use. However, the *manner of the subject's response has been more thoroughly standardized than the meanings of the situations to which he responds.* Other tests, such as the TAT, present somewhat more structured situations to the subject. However, the *clearer definition of the stimulus-material* tends to limit the extent of individual projection. The Symbol Elaboration Test, as we might call the application of the stimulus-patterns described herein, offers an advantage in that the situation presented to the subject is distinctly structuralized for the informed observer (clinical psychologist) while, for the subject, there is a minimum of conventional meaning to guide his responses. A further point in its favor is the fact that this test does not depend upon the subject's selection of certain elements of the field for his interpretation. In most techniques, it is sometimes a question of whether or not avoidance of a part of the stimulus is responsible for emphasis upon another part. Each of the symbols here presents a stimulus situation on an all-or-none basis. Thus, it is always clear to what situation specifically the subject reacts. Finally, the non-verbal character of the completion required in this test, prevents blockage in many subjects who might be inhibited on a verbal test having similar objectives, but who give no evidence of inhibition here.

Also important is the relative ease of interpretation made possible by the tester's previous knowledge of the implications of the symbols in this test. It is true that experience in free-drawing analysis is very helpful in interpreting the symbol-completions offered by the subject. However, even experience in handling symbolic material is not essential to the profitable use of the test proposed, so long as the symbols are accepted as carrying certain non-conscious implications. The inquiry-questions provide guidance in the interpretation of the drawings, such that the affectual tone of the subject toward a given symbol can be revealed. In every case there is some clue as to the subject's feeling toward the area of experience symbolized. Almost never does a subject totally refuse to draw upon a sheet; and, when he does, it may be discovered, or even assumed, that his conflict is with regard to the particular area symbolized on the sheet. The influence of the order of presentation of the symbols on the reactions of subjects has been proved elsewhere to be negligible.

The foregoing is not meant to imply that this test is expected to supplant the projective techniques already in use. Rather, it is felt that to propose the employment of a new test is to accept responsibility for justifying its addition to a standard battery. Anyone familiar with the various projective techniques now utilized in studying personality knows that each of them has a value of its own.

One important limitation to the use of this new test lies in the fact that only one main response can be given to each symbol in a single administration. The Rorschach test, the TAT, and others provide an opportunity for as many responses to a single stimulus as the subject wishes to make. Also, the fact that the symbols are centered on the page limits to an extent the applicability of the free-drawing type of analysis to this material. Another limitation lies in the effect of a strong dislike for drawing on the nature of the task. In such cases, the analysis depends solely on answers to the inquiry questions expressing the subjects' general attitude toward the stimulus material. Such an analysis is found to be schematic. Nonetheless, it is hoped that there will be further use for the symbols in clinical as well as research work.

BIBLIOGRAPHY

1. Alschuler, R. H., and Hattwick, L. W. *Painting and personality*. Chicago: Univ. of Chicago Press, 1947.
2. Barron, D. B. A study in symbolism. *Psychoanal. Rev.*, 1947, **39**, 395-431.
3. Brill, A. A. (Ed.) *The basic writings of Sigmund Freud*. New York: Random House, 1938.
4. Buck, J. N. The H-T-P technique. Qualitative and quantitative scoring manual. *J. clin. Psychol.*, 1948, **4**, 317-397.
5. Ferenczi, S. Zur Ontogenese der Symbole. *Int. Z. (ärtzl.) Psychoanal.*, 1913, **1**, 436-438.
6. Ford, C. S. Society, culture and the human organism. *J. genet. Psychol.*, 1939, **20**, 135-179.
7. Franck, K. Preferences for sex symbols and their personality correlates. *Genet. Psychol. Monogr.*, 1946, **33**, 73-123.
8. Gutheil, E. A. *The language of the dream*. New York: Macmillan, 1939.
9. Hallowell, A. I. Freudian symbolism in the dream of a Saulteaux Indian. *Man*, 1938, **38**, 47-48.
10. Horney, K. *New ways in psychoanalysis*. New York: Norton, 1939.
11. Immergluck, L., and Cohen, B. Motor responses to verbal stimuli. (Unpublished)
12. Jung, C. G. *Psychology of the unconscious*. (Trans. by B. M. Hinkel.) Boston: Moffat Yard, 1916.
13. Klopfer, B., and Kelley, D. M. *The Rorschach technique*. New York: World Book, 1942.
14. Lindner, R. M. Content analysis in Rorschach work. *Rorschach Res. Exch.*, 1946, 121-129.
15. Miller, J. G. *Unconsciousness*. New York: Wiley, 1942.
16. Morris, C. W. Relation of different concepts of the symbol. *J. Phil.*, 1927, **24**, 281-291.
17. Morgan, C. D., and Murray, H. A. The Thematic Apperception Test. *Arch. Neurol. Psychiat.*, 1935, **34**, 289-306.
18. Roheim, C. *The eternal ones of the dream*. New York: International Universities Press, 1945.
19. Sears, R. R. *Survey of objective studies of psychoanalytic concepts*. New York: Social Science Research Council, 1943.
20. Seligman, C. G. Anthropological perspective and psychological theory. *J. roy. Anthropol. Inst.*, 1933, 193-228.
21. Stekel, W. *Die Sprache des Traumes*. Wiesbaden: J. F. Bergmann, 1911.
22. Symons, N. J. A note on the formation of symbols. *Intern. J. Psychoanal.*, 1925, **6**, 440-443.
23. White, W. A. Symbolism. *Psychoanal. Rev.*, 1916, **3**, 1-25.

VOLUME 64
NUMBER 5

WHOLE NO. 311
1950

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

Changes in Responses to the Minnesota Multiphasic Inventory Following Certain Therapies

By

WILLIAM SCHOFIELD

Departments of Psychiatry and Psychology
University of Minnesota

Accepted for publication, October 7, 1949

Price $1.00

Published by

THE AMERICAN PSYCHOLOGICAL ASSOCIATION
1515 MASSACHUSETTS AVE. N.W., WASHINGTON 5, D.C

COPYRIGHT, 1950, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION

ACKNOWLEDGMENT

THE WRITER gratefully acknowledges the encouragement and critical suggestions of Professor Starke R. Hathaway.

He is especially indebted to his wife, Geraldine B. Schofield, for many hours of assistance in the tedious task of recording the basic data.

The writer also wishes to thank the staff of the University of Minnesota Hospitals, Psychopathic Unit, for generous cooperation with the problems of collecting pre- and post-therapy data.

WILLIAM SCHOFIELD

TABLE OF CONTENTS

ACKNOWLEDGMENT iii

I. INTRODUCTION 1

II. PURPOSE, SUBJECTS AND PROCEDURES 5

III. RESULTS 9

A. Mean Profiles for the Normal and Clinical Groups 9

B. Group Differences in the Frequency and Nature of Response Changes 15

C. Item Change in Relation to Score Change 19

D. Characteristics of the Most Stable Items and the Most Variable Items in Each Group 21

E. Investigation of a Therapy-Susceptible Scale 24

IV. SUMMARY AND CONCLUSIONS 28

APPENDIX 32

BIBLIOGRAPHY 33

CHANGES IN RESPONSES TO THE MINNESOTA MULTIPHASIC INVENTORY FOLLOWING CERTAIN THERAPIES[1]

CHAPTER I

INTRODUCTION

THREE INFLUENCES stimulated the present study: (*a*) the intensive demand for more and better psychiatric services, (*b*) the attempts to formulate new systems of psychotherapy, and (*c*) the very recent but increasing efforts to establish objective criteria for the evaluation of therapeutic endeavors. Specifically, this study is concerned with an exploration of certain of the theoretical and practical problems involved in the measurement by means of a structured personality inventory of therapeutically induced changes in personality.

Psychiatry is currently experiencing a period of intense activity which is in direct response to the demands for therapy of the numberless personality disorders of all degrees engendered in civilian and veteran populations alike by the emotional stresses of World War II (1). With the addition of the personal maladjustments originating under the pressure of the emergency years to the pre-war case load, it is natural that the interest and contributions of the government and the public should be in the direction of increasing the number and quality of centers for the training of more clinicians. The approximately 4000 psychiatrists and 800 clinical psychologists now listed on the rolls of their respective organizations cannot, in point of numbers, handle in anything approaching adequate fashion the prophylactic and therapeutic needs of the population. Estimates have placed the number of psychiatrists required in certain areas at from four to seven times the number now available (5). It was primarily to remedy the dearth of adequately trained personnel that the National Mental Health Act was passed in 1946. Under the provisions of this act, the United States Public Health Service is directing and subsidizing an extensive program to improve training facilities in psychiatry, clinical psychology, and psychiatric social work and to increase the number of trainees in these fields.

It is necessary, though difficult, in the midst of practical considerations and the exigencies of daily service to attempt a long view. Some workers already have expressed doubt as to the ultimate validity of the present emphasis on training. Kubie suggests that there is a fallacy in the theory that the training of more workers will suffice to improve the psychiatric health of the community or nation, at least as measured in terms of the number of patients (12). He points out that increased availability of services invariably leads to increased demand, so that the "catching up" point is never reached. An obviously crucial factor in determining the patients-per-therapist ratio is the effectiveness of the remedies available to the clinician. Not only must more workers be trained, but their training and therapeutic armamentarium must be enhanced by new techniques of proven efficacy. The development and testing of new techniques of therapy demand a research orientation, and research programs must not be grossly subordinated to the construction of curricula if psychiatry is to derive maximum benefit from the advantage it now enjoys under the triple incentives of social need, favorable public opinion, and government funds.

It is fortunate that the effective relationships which were established between psychiatry and clinical psychology in the armed services show promise of being maintained. Clinical psychology also is benefiting from government subsidization and while this support is, again, primarily to increase the number of trainees in this field, it may be hoped that a direct impetus to research will result.

It is only through the medium of research, supported and conducted by all professional workers concerned with the treatment and prevention of psychiatric disorders, that we may hope to achieve the construction and sharpening of those diagnostic, prognostic, and evaluative instruments which are necessary to assure the most efficient application of such facilities as are available for the care of personality disturbances at all levels of their manifestations.

---

[1] This paper is a part of a thesis submitted to the Department of Psychology of the University of Minnesota in partial fulfillment of the requirements for the degree of Doctor of Philosophy, June, 1948.

What are the most pressing research needs in the field of mental health from a long-term view? First of all, there is a need for research which will establish in a definitive manner the level of effectiveness of existent therapies with respect to specifically diagnosed disorders. Secondly, there is a need for research which will establish the validity of newly proposed therapies. Inherent in researches of this type is the specific problem of determining which patients improve, and how much they improve, and which patients do not improve under a given regimen. The solution of this problem is contingent upon the attainment of a standard, objective, representative evaluation of the status of an individual's personality.

This paper presents an investigation of some of the problems arising in the application and interpretation of such evaluations, particularly as they furnish data on the effectiveness of given therapies.

The histories of the major psychiatric therapies have followed a certain pattern. In brief, there has been an initial enthusiastic announcement of the new technique accompanied with or followed shortly by figures on recovery which have represented not merely a sizeable improvement over those for earlier therapies but which have approached the 100 per cent point so closely as to suggest a specific cure. Later reports on the technique from other clinics and sometimes from the clinic of origin have shown considerably lower frequencies of recovery. Thus, Sakel in first publishing on insulin shock therapy of schizophrenia claimed 87.9 per cent remissions. This figure has not been duplicated by any other workers.

There are undoubtedly many factors involved in the explanation of divergences between comparable figures for recovery rates in two clinics as well as for the failure of originally very high rates of remission in a given clinic to be maintained in further studies. A considerable part of divergence in interclinic statistics is probably due to differences in the systems (or lack of system) by which the status of patients is evaluated both before and after a course of therapy. Most commonly such evaluations are characterized by subjectivity, with little concern that the evaluation of the patient before and after therapy be on the same schedule of variables (3). In extreme form, this type of evaluation of a patient's progress is expressed in terms of the facts that his diagnosis at time of admission to the hospital was "schizophrenia—paranoid type" and that he was noted at time of discharge to be free of hallucinations. There would be many observations and facts concerning the patient which went "into" the diagnosis and yet which, in the form of a diagnosis, would not be available as a basis for evaluating his status when discharged. Likewise, there could be many observations concerning the patient at the time of discharge which would be relevant to predictions of his adjustment at home and at work and yet which would not be properly weighed and communicated by recording only that his hallucinatory behavior had ceased. *Evaluations of this gross type permit recognition of the most obvious degrees of improvement, i.e., abatement of severe symptomatology, but they are not sufficient for detection of improvement in the patient's personality with respect to those inadequacies in social adjustments and diurnal habits which served as the first clues to his illness.* From this point of view, the maximal efficacy of a given therapy for a given patient could not be ascertained at the time of his discharge from the hospital but only after he had returned to a normal, non-institutional environment and only in terms of the quality of his adjustments in that environment. Furthermore, the observations and evaluations of his post-therapy behavior would have to be ordered to the same method of recording as was used to depict his status prior to therapy. Limitations of time and money have generally prohibited the attainment of this ideal paradigm for the validation of a therapeutic procedure. The next best approach demands that the status of the individual be thoroughly measured in terms of such observations as are possible immediately prior to the start of a course of therapy and immediately upon the completion of that course. It is equally important in this more circumscribed evaluation of the patient's progress that his status before and after therapy he measured with respect to the same variables.

It may be argued that no matter how rigorous the procedure of investigation, the "pure" effects of shock therapy, for example, may not be determined since such therapy is always imbedded in a certain setting and adminstered by personnel whose effects on the patients cannot be held constant. It may further be argued that the effects assignable to shock are inextricably mixed with the effects of the less dramatic psychotherapeutic efforts of the physician. With regard to this last point, it must be recognized that the ancillary therapeutic programs developed for the patient receiving convulsive therapy are probably as diverse as the theories proposed to explain the mechanics (or dynamics) of shock as therapy. This source of discrepancy should be remembered when it is necessary to explain different measured effects in different clinics when the same (?) therapy has been applied.

With regard to the complicating effects of other psychotherapeutic efforts (e.g., interviews) accompanying a major medical therapy, it should be recognized that the clinic in which the decision to apply a convulsive therapy is *not* accompanied by a reduction (either planned or unplanned) in the number of hours spent by the physician in personal contacts with the patient is probably rare. When a gross procedure of the nature of electroshock seems a promising approach with a given patient, the therapist probably feels some justification in reducing his contacts with that patient so as to afford more hours to those patients for whom no "specific" therapy is available. In short, the effects of a therapy such as electroshock are probably open to evaluation in a more pure form than would at first appear to be the case.

The design and carrying through of investigations of the above nature must necessarily await the development of adequate techniques for the thorough descriptive assay of the individual personality. It has been only within the last decade that clinical psychology has offered an instrument having the essential virtues of psychiatrically oriented variables empirically validated and quantitatively expressed. This instrument is the Minnesota Multiphasic Personality Inventory (8, 9).[2] This psychiatric inventory was designed primarily to furnish an external, objective source of information bearing on the detection and diagnosis of personality aberrations. In addition to four scales which indicate the validity of the subject's record, the MMPI measures personality deviations on nine clinical scales: Hs–hypochondriasis, D–depression, Hy—hysteria, Pd—psychopathic deviate, Mf—masculinity-femininity, Pa—paranoia, Pt—psychasthenia, Sc—schizophrenia, and Ma—hypomania.

---

[2] The Minnesota Multiphasic Personality Inventory will be referred to hereafter as the MMPI.

The validity measures are the "Cannot Say," L, F, and K scales. The first of these is simply the number of items placed in the "Cannot Say" category. Since the "Cannot Say" responses are not scored on any of the other scales, a large number of such responses has a generally suppressing effect on the clinical profile. The L scale is composed of fifteen items designed to measure the strength of the subject's tendency to "fake good" by selecting responses which place him in a socially acceptable light. The F scale is composed of items the scored responses to which are extremely infrequent; elevated F-scores may occur as a result of errors in scoring, carelessness or lack of comprehension by the subject, or a "fake bad" tendency in the subject. The K scale is the most recent of the validity scales and provides a measure of the degree to which "defensiveness" has characterized the subject's test-taking attitude.

The wide use of the MMPI by psychiatrists, clinics, and general physicians attests to its success in a diagnostic role (2, 15). It is interesting to note that only recently has there been a sign of appreciation of its potential value in measuring the effects of a given course of therapy (7, 15, 16). That there should be such a lag in the use of a basically diagnostic instrument for the evaluation of change reflects a general failure on the part of psychiatrists and psychologists to recognize that the determination of the patient's condition at any time after hospitalization or the onset of therapeutic interviews cannot be usefully made except on the very same variables the assessment of which served as the basis for hospitalization or the onset of therapy. The clinician should be no less willing to accept supportive information for his post-therapy evaluation than he was to seek assistance in the original problem of diagnosis.

CHAPTER II

PURPOSE, SUBJECTS AND PROCEDURES

THE MAJOR ASSUMPTION underlying the investigation reported here is that an instrument of demonstrated validity in the measurement of personality deviations of psychiatric import is of equal validity for the evaluation of personality changes and of the therapeutic operations deemed responsible for those changes. It is not assumed, however, that a direct evaluation of change in personality, or of the efficacy of a therapy, is possible through the single study of score changes. Rather, it is believed necessary to determine whether there are patterns of change in scores on a multivariable instrument which are characteristic for given patients under given regimens and to discriminate, on as objective a basis as possible, between those changes in the personality assay which are therapy-related and those which are merely chance. Further, it is believed necessary to evaluate the extent of the relationship between changes in item response and associated score changes.

In the event that an individual completes a personality questionnaire such as the MMPI on two different occasions, certain analyses may be made of any differences which are observed in the resulting profiles. It may be asked if the two profiles differ in a degree which would appear to approach significance. More specifically, this question relates to whether the two scores obtained on each single scale of the profiles differ by such an amount as exceeds that for which the hypothesis of no true difference (between the means of hypothetical populations of first-profile and second-profile scores) is tenable. For simplicity in reference, this is the problem of *status-score change*. The observed differences between the various pairs of status-scores constitute *change-scores.*

There is still another kind of change not yet extensively considered, perhaps because attention to status-score changes has tended to mask the operation of this more subtle, but potentially significant variation from profile to profile. This is the change in response to specific items composing the various scales; it may be referred to as *item-response change.*

The purpose of this investigation has been to study the item-responses on the MMPI of certain groups of persons and the changes in the item-responses of these persons as a result of the passage of time plus the occurrence of certain events. More specifically, it has been designed to study the nature of changes in the responses to the MMPI items of certain groups of psychiatric patients and the relationships between item-response changes and status-score changes in these groups, and it has been oriented toward the problem of detecting and measuring therapy-induced changes in personality.

The investigation began with a careful check of the approximately 7000 records in the MMPI research files at the University of Minnesota to obtain the names of all persons, normals and patients, to whom the inventory had been administered on at least two occasions. For all cases, available information on education, occupation, marital status, place of residence, etc. was recorded in addition to the interval, to the nearest 5-day unit, between the two administrations of the MMPI. No in-patients were included in the study unless a check of their hospital record revealed that they had been discharged "improved" or "recovered" and that the two MMPI profiles had been obtained within the limits of a single hospitalization, and the length of this hospitalization was recorded in addition to the above information.[1] The hospital charts of the patients were also checked carefully to determine the exact nature of the therapy which the patient had received, and particular note was made of the use of electroshock.

A post-admission MMPI is obtained routinely on all patients admitted to the Psychopathic Unit of the University of Minnesota Hospitals. With the beginning of this study in September, 1946, a special effort was made to collect post-therapy, pre-discharge MMPI records on all patients. Most of the patients comprising the "psychotic-shock" group for this study were obtained in this manner.

The Psychopathic Unit has a total of only 35 beds of which six are in a separate ward for grossly disturbed patients. While the number of beds available to men and women patients on the general psychiatric ward are approximately equal, the actual number of women hospitalized per year greatly exceeds the number of men.[2] Due to lack of an adequate number of male patients for whom two records are available, the present study deals entirely with females.

The normal cases fall into two groups: the first is composed of nineteen female night-school students, half of whom took the individual form of the MMPI first and one week later took the group form, the other half taking the two forms in reversed order with a one-week interval between them.[3] The second group is composed of 23 females selected from the files because they had two MMPI records. The mean interval between the two tests for this second group was 136.4 days; the median interval was 47.5 days. This indicates the marked positive skewness which characterized the distribution of inter-test intervals. The mean age of the first group was 32.1 years, with a standard deviation of 7.3; the mean age of the second group was 26.3 years, with a standard deviation of 4.4. The difference in the mean ages of the two groups is not significant ($t = .998$) nor is the difference in the variances significant ($F = 2.727$). The MMPI test and retest means and standard deviations for the two normal groups are given in Table 1. Only one of the test-retest differences in mean score is statistically significant; this difference occurs on the Pt scale in the night-school sample ($t = 3.35$; $P < .01$). The test-retest correlation for the Pt scores in this group is .783, the highest of such correlations in this group for any of the clinical scales. If we may assume, on the basis of the very marked tendency in both groups of normals for the second test means on the clinical scales to be lower than the original means and on the basis of similar findings in other studies which have indicated a tendency for the "adjustment scores" of normals to change in the direction of improvement, that there exists an

---

[1] In the case of the out-patients included in this study, the number of visits to the Out-patient Psychiatric Clinic between the two MMPIs was recorded. No evaluation of "improvement" in these patients was available.

[2] In 1947, the Psychopathic Unit treated a total of 180 female and 123 male in-patients.

[3] These cases were originally included in a study to determine the equality of the group and individual forms. No significant differences were found between the mean scale scores obtained with the two forms. The writer is indebted to Dr. Howard P. Longstaff for the data from these cases.

TABLE 1

TEST AND RETEST MEANS AND STANDARD DEVIATIONS OF THE MMPI SCORES OF THE TWO GROUPS OF NORMALS*

Group	Test	? M	? σ	L M	L σ	F M	F σ	K M	K σ	Hs M	Hs σ	D M	D σ
Night-school	1	12.6	9.7	3.2	2.2	3.4	3.1	55.2	8.6	48.0	5.4	51.4	8.8
	2	17.3	19.0	3.4	2.2	2.6	3.2	56.3	8.6	46.0	5.7	49.0	7.3
File-random	1	20.4	17.8	3.4	2.3	3.0	2.0	57.9	7.8	46.3	7.8	47.0	10.0
	2	21.9	20.4	3.6	2.0	2.4	1.8	59.0	7.8	45.4	5.7	45.3	7.6
		Hy		Pd		Pa		Pt		Sc		Ma	
Night-school	1	52.1	6.8	54.0	8.8	53.8	10.2	52.1	6.6	53.1	7.5	51.4	8.2
	2	50.9	7.8	52.9	9.2	50.6	9.3	48.7	6.8	51.5	8.4	49.0	8.7
File-random	1	50.4	8.6	49.6	7.1	50.2	8.6	47.8	7.2	52.2	7.8	48.4	9.4
	2	48.2	8.4	49.2	8.5	51.0	7.0	47.2	7.8	50.7	6.2	49.2	10.8

[*] The MMPI data in this and all subsequent tables are based on T-scores derived from the usual formula:

$$T = 50 + 10\frac{X - M}{\sigma}$$

except for the data for the "?," L, and F scales which are in raw score units.

as yet unanalyzed factor making for a regression toward the norm in the personality measure of normals, even in the absence of originally extreme scores, then the appearance of such a factor would be maximized in the test-retest differences of those scales having the greatest reliability. This explanation of the change on the Pt scale shown by the night-school students is offered in the absence of evidence of any more obvious factor which might account for the observed difference. It will be noted from Table 2, which presents a summary of descriptive statistics for the normal and clinical groups, that the combined normal group is characterized by a high incidence of individuals with high school

TABLE 2

GENERAL DESCRIPTIVE STATISTICS FOR THE FIVE SAMPLES TESTED

Group[a]	Number	Age Mean	Age Range	Days in Hospital[b] Mean	Days in Hospital[b] Range	Days between MMPIs Mean	Days between MMPIs Range	% Single	% H.S.[c]
Normals	42	28.9	21–46			61.6	2–1060	78.6	95.0
OPD	25	32.1	17–56	5.8	2–18	80.6	5–330	32.0	63.2
PN	24	26.9	17–39	52.0	20–114	33.4	5–75	41.7	34.8
PS-NS	13	35.8	20–57	64.4	15–106	33.8	5–65	38.5	30.8
PS-S	20	43.2	20–59	50.6	23–90	27.2	5–40	30.8	83.3

[a] OPD—25 out-patients.
PN—24 hospitalized neurotics.
PS-NS—13 hospitalized psychotics who received neither insulin nor electroshock therapy.
PS-S—20 hospitalized psychotics treated with electroshock.

[b] The figures in this column for the OPD group refer to number of clinic visits in the interval between the two MMPIs.

[c] Percent of group with high school education or better.

education or better, from metropolitan areas.

The clinical groups are four in number: 25 out-patients; 24 hospitalized neurotics; 13 hospitalized psychotics who did not receive any type of convulsive therapy; and 20 psychotics who received electroshock therapy. Table 3 presents the composition of the clinical groups by diagnosis.

TABLE 3

THE COMPOSITION OF THE CLINICAL GROUPS BY DIAGNOSIS

Group	Diagnosis	Number
OPD N = 25	Psychoneurosis (mostly mixed)	22
	Schizophrenia, simple	1
	Hypomania	1
	Undiagnosed	1
PN N = 24	Psychoneurosis, mixed	13
	hysteria	4
	hypochondriasis	2
	psychasthenia	1
	reactive depression	4
PS-NS N = 13	Manic-depressive, depressed	3
	manic	2
	circular	1
	Agitated depression	1
	Schizophrenia, simple	1
	catatonic	1
	Recurrent depression	1
	Psychosis, undiagnosed	3
PS-S N = 20	Manic-depressive, depressed	4
	Involutional, melancholia	8
	paranoia	2
	Schizophrenia, paranoid	2
	hebephrenia	1
	undetermined	1
	Paranoid condition	2

The two MMPI records of each subject were tallied on a common form in such a way as to indicate the actual response ("True," "False," "Cannot Say") of the subject to each item on each of the two testings. From these combined records, it was possible to determine for each individual the total number of changed responses, the number of changes by direction of change, and the number of changes per general type of item. These data for the individuals were then combined to permit determination of group statistics.

The mean score of each group on each of the validity and clinical scales of the MMPI was determined for the first and second tests and the reliability of the differences between the test and retest means of each group was determined by statistical analysis.

The inter-group differences in the means and variances for the number of item-responses changed per individual were tested for reliability as were the inter-group differences in means and variances for the number of responses changed per item.

In order to study the relationship of item-response changes to changes in status-scores, the total change-score, i.e., change in status-scores, for each individual was determined by adding, without regard to sign, the differences between his test and retest scores on the clinical variables. The ratio of this total change-score (Σcs) to the number of items changed (Σic) was determined for each individual and summary statistics for this ratio (Σcs/Σic) were determined for each of the groups.

Finally, item-by-item comparisons were made to determine the reliability of the differences between the per cent of responses changed by the normal group and each of the clinical groups respectively. Lists of items with reliable differences in frequency of change were compiled and studied to determine if they showed any dominant characteristics.

In view of the small *N* for each of the patient groups, the present study can only be considered as exploratory rather than definitive, a search for support or refutation of specific hypotheses, rather than a basis for final establishment of conclusions.

Chapter III

RESULTS

A. Mean Profiles for the Normal and Clinical Groups

The means and standard deviations of each of the groups on each of the validity and clinical scales of the MMPI, for both administrations of the inventory, are given in Table 4. The corresponding profiles are presented in Figures 1-5.

The significance of the difference between the first and second administration means for a given scale in a given group was tested by the following formula:

$$t_{(M_1-M_2)} = \frac{M_1 - {}_2M}{\sqrt{\dfrac{N\Sigma D^2 - (\Sigma D)^2}{N^2(N-1)}}}$$

where D = the difference between the first and second scores of a given individual. Only those differences have been considered reliable which yield a *"t"* score with a probability of less than 1 per cent.

It will be noted with respect to the clinical variables that there is a general tendency for the mean scores on the second administration to be lower than those for the first. Exceptions to this occur on the Pd scale in the out-patient (OPD) group, on the Mf scale in the hospitalized neurotic (PN) group, and on the D and Hy scales in the psychotic patients who did not receive electroshock therapy (PS-NS). None of the mean score differences of the combined Normal Group is statistically significant nor is any of the OPD group. The mean F-score difference in the Normal Group has a P of $< .02$.

In the PN group, the second test means on the Hs, Hy, and Pd scales are significantly lower than the original means ($P < .01$ in each case).

The PS-NS group shows no significant mean differences on any of the clinical scales. The difference between the two K-score means is significant at the 1 per cent level.

The sample of psychotic patients who received electroshock therapy (PS-S) shows a reliable decrease of mean score on the Hs, D, Hy, Pd, Pt, and Sc scales. The mean difference on the Pa scales has a probability of less than 2 per cent.

With respect to the validity scales, it should be noted that there is a consistent tendency for the F-score mean to be smaller on the second test, while the reverse holds for the K-score except for the PS-S group. None of the mean differences for these two variables is reliable, however.

From Table 2, it may be seen that the Normal Group is characterized by a higher proportion of single persons and persons with a high school education or better than characterizes the clinical groups. From Fig. 1 and Table 4, it appears that these selective factors have not unduly influenced the "normality" of this group insofar as the clinical scales of the MMPI are concerned. The two profiles follow the 50-score line closely and there is not the elevation on the Hy-scale to be expected in a college group. Furthermore, the test-retest reliabilities for the various scales in the combined normal group closely approximate those reported for normals in the literature.

In working with personality scores in

TABLE 4

TEST-RETEST MEANS AND STANDARD DEVIATIONS FOR THE SCORES OF EACH GROUP ON THE MMPI, WITH SIGNIFICANCE TESTS FOR THE MEAN DIFFERENCES[a]

Group	Test	?		L		F		K		Hs		D	
		M	σ	M	σ	M	σ	M	σ	M	σ	M	σ
Normals *N* = 42	1 2	17.0 19.8	15.2 19.9	3.2 3.4	2.2 2.1	3.2 2.6	2.6 2.6	56.6 57.8	8.3 8.3	47.1 45.7	6.9 5.7	49.0 47.0	9.8 7.7
OPD *N* = 25	1 2	14.8 8.0	19.0 8.0	4.4 4.2	2.2 1.6	7.0 6.2	4.0 3.4	51.3 52.4	9.0 8.2	74.4 70.1	13.4 12.6	73.5 72.7	13.4 9.0
PN *N* = 24	1 2	14.2 11.6	18.6 12.4	4.6 4.2	2.6 2.4	6.1 5.2	3.9 5.0	52.2 53.4	9.8 8.5	74.9 66.5	13.4 12.4	71.8 67.6	8.3 9.7
PS-NS *N* = 13	1 2	27.4 11.4	23.4 9.3	4.3 4.5	2.2 1.3	10.9 6.6	9.4 6.7	44.6 52.5	6.9 6.8	58.6 57.8	13.3 11.2	64.9 66.7	11.0 7.3
PS-S *N* = 20	1 2	14.4 23.0	13.6 23.4	4.6 3.7	2.4 2.0	8.2 4.0	6.0 2.5	52.1 50.7	7.2 6.9	62.8 53.0	10.7 12.8	74.4 58.2	10.8 9.1
		"*t*"		"*t*"		"*t*"		"*t*"		"*t*"		"*t*"	
Normals		1.4		.8		2.1		1.0		1.6		2.1	
OPD		2.0		.8		1.6		.7		2.0		.3	
PN		1.4		1.0		.5		.6		3.2*		1.4	
PS-NS		2.4		1.6		1.3		3.0*		.2		.5	
PS-S		1.7		1.8		3.2*		1.0		4.0*		4.2*	

The data for the "?," L, and F scales are in raw score units. Those for the remaining scales are in T-score units.

[b] The *N*'s on which the means and standard deviations for the Mf score are based are: OPD = 20; PN = 13; PS-S = 20.

* Significant at the 1% level.

which deviation is permitted in one direction only, in which experience with like measures indicates a general phenomenon of regression in the direction of "normality" (i.e., lower scores) with a simple passage of time, and in which the expected direction of change for patients receiving specialized therapy is toward lower scores on a retest, we would expect the retest scores of patients to show a reduced range and to find this reduced range reflected in lower test-retest correlations than are found with normals tested over a similar interval. In general these expectations are borne out by the results in Table 5, which show a tendency for the test-retest correlations to be lower for the patient groups than for the normals. In view of this argument, it may be considered more proper to consider test-retest correlations in patients who have had therapy as indices, at least in part, of therapeutic efficacy rather than as reliability coefficients. The small size of the samples in this study precludes testing the significance of the differences between these coefficients but their values are reflected in the tests of the reliability of the differences in corresponding means since they enter into the determination of the standard error of the difference between means for the test and retest scores. In this respect, it should be noted that Darley (4) interpreted his test-retest coefficients, which were considerably lower than the corrected

TABLE 5

TEST-RETEST CORRELATIONS OF EACH GROUP ON EACH SCALE

Group	?	L	F	K	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma
Normals (N = 42)	.718	.810	.749	.663	.724	.734	.682	.569		.585	.701	.594	.783
OPD (N = 25)	.430	.556	.797	.580	.657	.424	.664	.568	.827	.367	.708	.796	.746
PN (N = 24)	.884	.596	.687	.763	.536	.570	.617	.605	.589	.498	.475	.608	.819
PS-NS* (N = 13)	.334	.852	.375	.105	.220	.707	.436	.375		.052	.147	−.303	.608
PS-S (N = 20)	.434	.555	.384	.353	.908	.486	.555	.431	.343	.580	.496	.660	.456

* The coefficients for the PS-NS group are product-moment "r's" corresponding to computed "rho's" as given in Table 45, p. 362, Garrett, *Statistics in Psychology and Education*, 2nd edition.

TABLE 4—*Continued*

Hy		Pd		Mf^b		Pa		Pt		Sc		Ma	
M	σ	M	σ	M	σ	M	σ	M	σ	M	σ	M	σ
51.6	8.0	51.6	8.2			51.8	9.5	49.8	7.2	52.6	7.6	49.8	9.0
49.4	8.2	50.8	9.0			50.8	8.2	47.9	7.4	51.0	7.2	49.1	9.8
76.2	11.2	61.3	12.2	48.8	12.1	61.2	10.6	69.5	11.4	69.8	12.8	55.9	11.9
73.1	12.4	66.2	14.8	48.8	9.8	59.7	9.4	68.4	13.6	67.3	13.2	55.0	10.5
77.2	10.8	63.6	13.2	42.8	8.9	62.5	10.6	66.6	11.8	69.2	12.4	55.2	12.7
69.2	11.3	60.5	11.0	46.2	10.6	59.1	10.2	65.8	12.7	67.0	14.4	55.8	12.6
62.0	12.0	58.8	14.4			70.0	12.3	66.2	11.4	70.0	16.3	59.0	18.0
63.1	10.4	58.1	6.8			66.0	14.4	63.4	8.6	64.8	15.6	53.3	17.8
65.2	13.5	72.0	15.8	49.4	10.0	70.2	14.2	70.9	17.2	72.2	19.7	56.2	12.2
55.7	12.2	57.6	13.3	49.4	9.2	59.2	19.4	57.6	15.3	59.8	15.7	52.5	12.3
"*t*"		"*t*"		"*t*"		"*t*"		"*t*"		"*t*"		"*t*"	
2.0		.6				.8		2.0		1.4		.6	
1.6		1.8		.1		.8		.5		1.4		.6	
4.0		4.2		1.4		1.5		.2		.8		.4	
.2		.2				.6		.8		.7		1.2	
3.4		4.2		1.6		2.6		3.5*		3.6*		1.2	

odd-even reliabilities for his measures, as guages of "trait stability," representing a non-reducible combination of actual change in his subjects plus errors of measurement.

The similarity of the profiles for the OPD and PN groups (Figures 2 and 3) is notable. The pre- and post-therapy profiles for both of these groups show the general form which has been described as "typical" for neurotics (6). In terms of such typical profiles, the elevation on the neurotic triad for these groups would suggest that they were composed of so-called "mild" psychoneurotics. The height of the secondary elevation on the Pa, Pt and Sc scales, however, is greater than that to be expected in mild psychoneurotics and suggests that both the out-patient and hospitalized groups were composed of rather severe problems. This might explain at least partially the ineffectiveness of the out-patient therapy in reducing the neurotic triad although the relatively small number of therapy sessions intervening between the two profiles (see Table 2) is undoubtedly a factor also.

"Therapy" as provided by the OPD clinic requires definition. The patients have their most frequent and extensive contacts (for history-taking, therapeutic interviews, etc.) with "junior clerks." These are junior medical students who are assigned to the psychiatric service for a ten-week period. In their OPD activity, they work in close alliance with the psychiatric residents and are supervised by the senior staff psychiatrists and clinical associates. While the patients are seen regularly (weekly or bi-weekly) by the junior clerks, they are seen only occasionally by a staff psychiatrist.

The rise in the mean Pd score of the OPD group, contrasted with the significant decrease in the same score for the hospitalized neurotics, presents an intriguing problem of explanation. One hypothesis which might explain these relationships is based on the relative "social acceptance" of the neurotic symptomatology experienced by the out-patient versus the in-patient. In the case of the out-patient, the probing of the therapist to uncover possible conflictual material which might have an etiologic

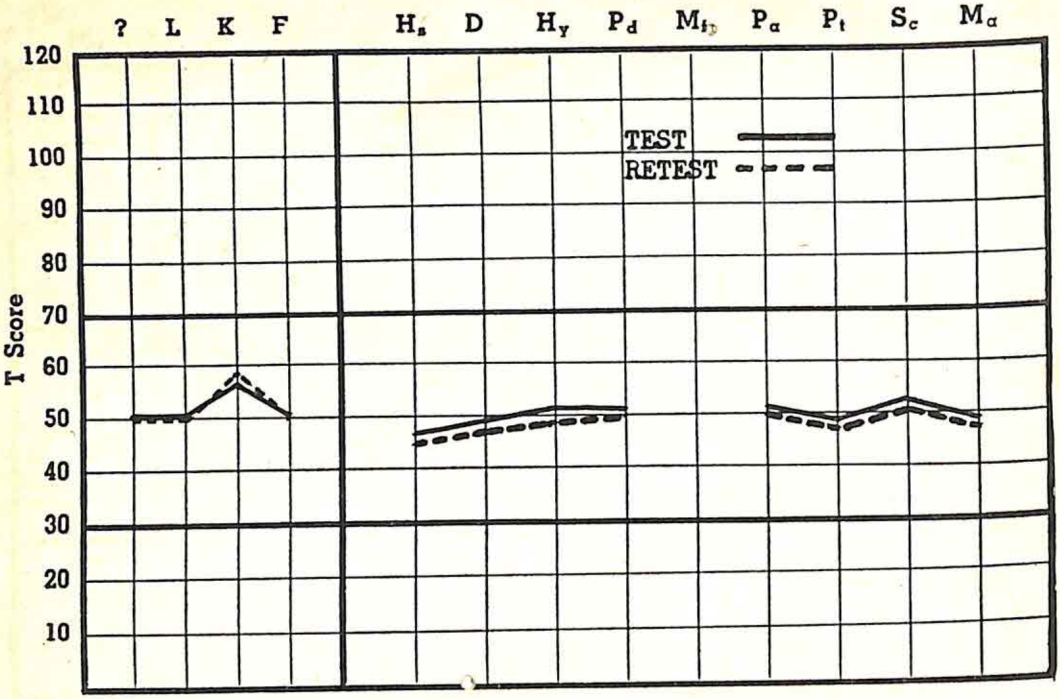


FIGURE 1. Mean test and retest profiles of the combined normals. ($N = 42$.)

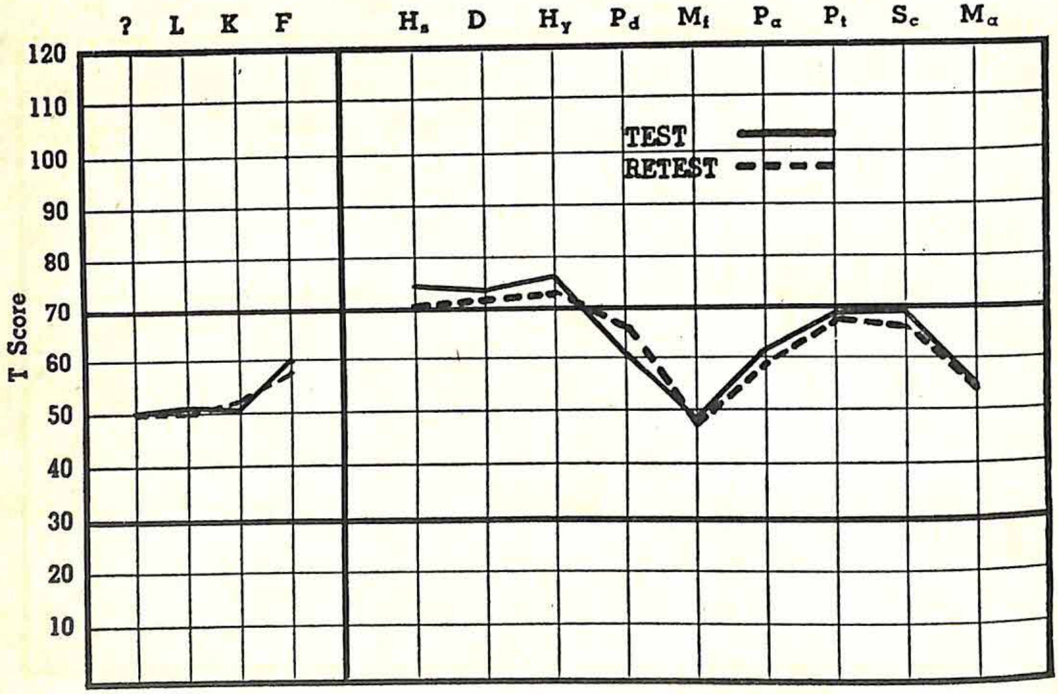


FIGURE 2. Mean pre-therapy and post-therapy profiles of the out-patients. ($N = 25$.)

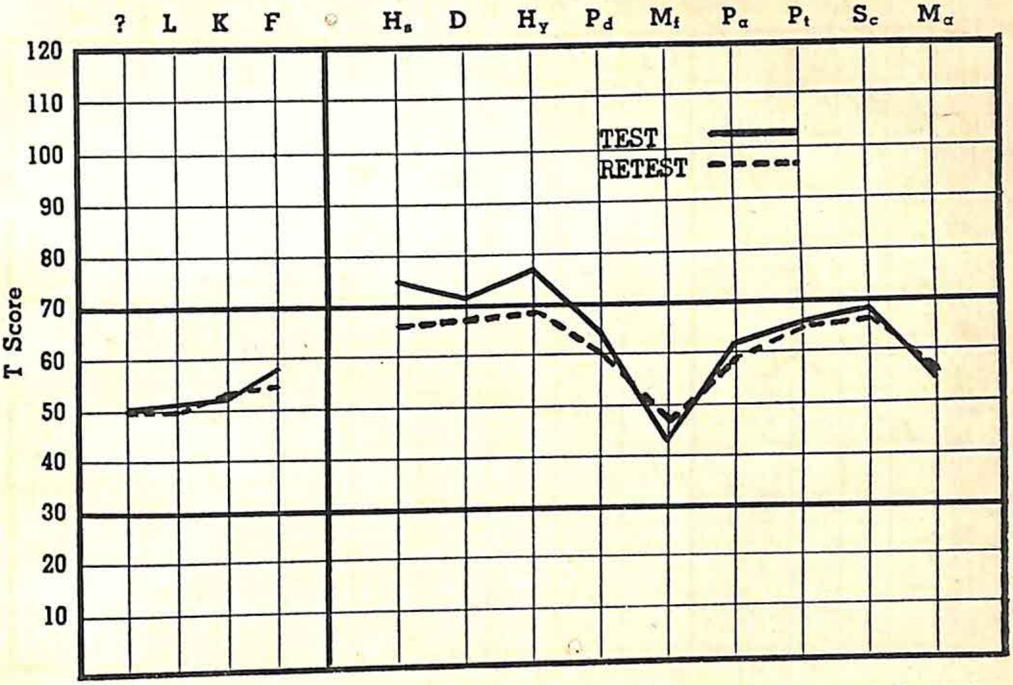


FIGURE 3. Mean pre-therapy and post-therapy profiles of the hospitalized neurotics. ($N = 24$.)

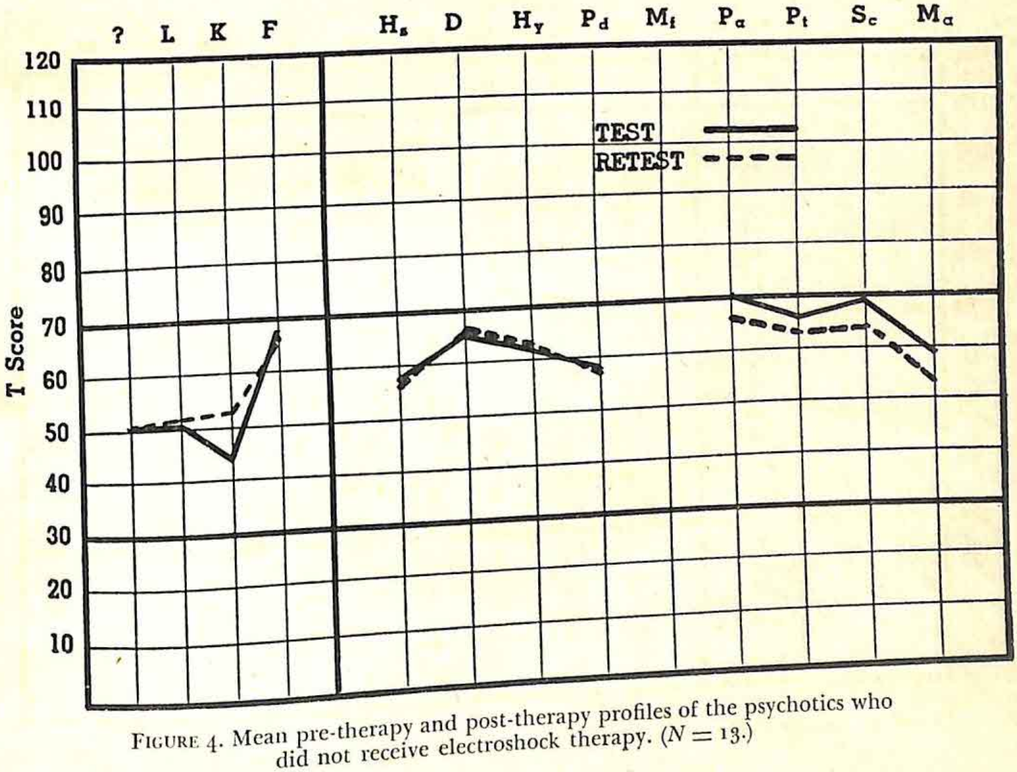


FIGURE 4. Mean pre-therapy and post-therapy profiles of the psychotics who did not receive electroshock therapy. ($N = 13$.)

role and his apparent discounting of the gross symptomatology may result in a sensitizing of the patient to the quality of his social adjustments to the extent that, with later administrations of the MMPI, he becomes a "plus-getter" on those items of the Pd-scale which pertain to family and social adjustments, while his neurotic symptoms are relatively unaffected.[1] The lower mean scores on the L and Hy scales for the second administration of the MMPI to the OPD group, while not significantly different from the original means, are in accord with this hypothesis.

In the case of the hospitalized neurotic both the severity of symptoms which justified intramural therapy and the fact that hospitalization occurred, mean that the therapist and patient alike may legitimately note the progress of therapy in terms of the amelioration of gross symptomatology and there is less of the subtle pressure on the patient to continue to uncover interpersonal conflicts.

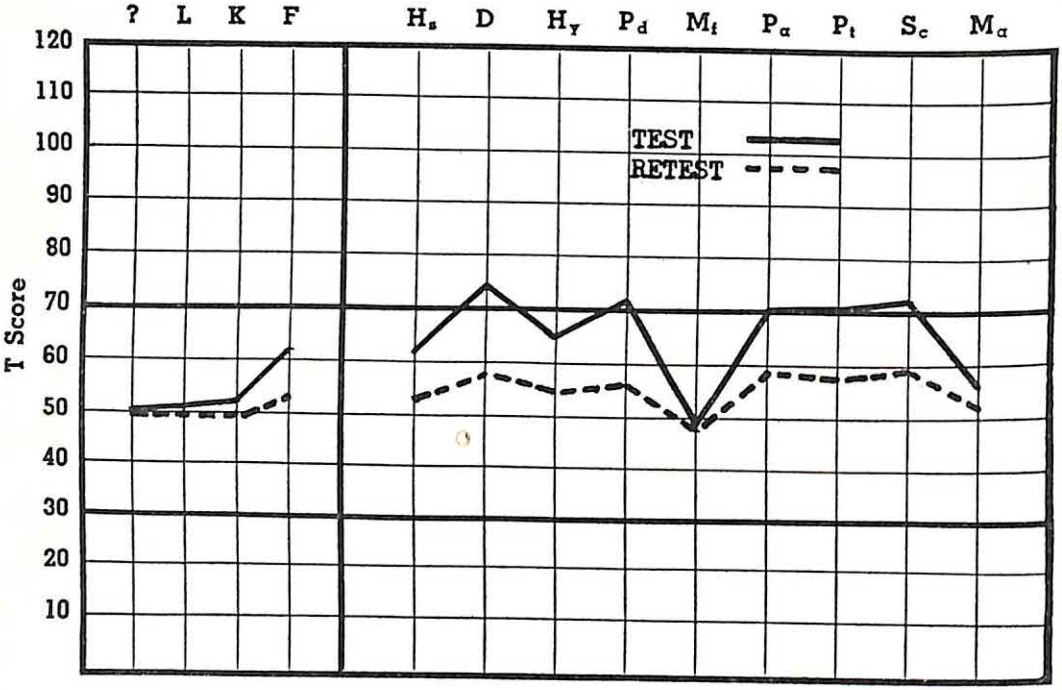


FIGURE 5. Mean pre-therapy and post-therapy profiles of the psychotics who received electroshock therapy. ($N = 20$.)

This hypothesis is supported somewhat by the data from Table 9 which shows the percentage of item responses changed for those items classified as pertaining to "family and marital" concerns (Category XI) to be less for the PN than for the OPD group. Conversely, the OPD group shows a tendency to smaller percentages of item-responses changed than the PN group for those items classified as pertaining to general neurologic, cranial nerves, motility and coordination, and sensibility (Categories II, III, IV and V).

It is interesting to note that with psychotic patients, the ordinary regimen of the psychiatric ward in the absence of electroconvulsive therapy is not adequate to achieve a significant reduction of the mean scores on the psychotic scales. While all of the psychotic scales in Figure 4 show reductions in mean score, there is a slight though not significant increase in the mean scores for the D and Hy scales. The relative "mildness" of the profiles plus the fact that these patients did not receive electroshock suggests that they were borderline or mild psychotics with extensive neurotic coloring and this fact may account for the apparent shift from psychotic to neurotic symptomatology. Concomitant with this shift, there is an increase in defensiveness or control as revealed in a mean K-score increase of 7.90 ($t = 3.000$, $P < .01$).

---

[1] The term "plus-getter" has been applied to those individuals with a specific motivation, conscious or otherwise, to put themselves in an unfavorable or socially undesirable light.

Figure 5, which presents the two mean profiles for the group of psychotic patients who received electroconvulsive therapy is of interest not only because of the marked changes it reveals but because of the stability shown by the Mf and Ma scales. The stability of these two scales in a group which shows so much change on the other clinical variables is of great importance as evidence for the selective sensitivity of the MMPI in reflecting therapy-induced changes. Had all of the MMPI scales shown significantly reduced means following electroshock therapy, it might be argued that all that was being measured was a gross tendency to response-change as a result of the confusion and memory defects associated with EST and that specific personality changes were masked. The differential change reflected in the profiles of Figure 5 argues for the efficiency of the MMPI as an instrument for detecting the particular results of special therapies.

In summary, the data on the mean pre- and post-therapy profiles reveals no significant changes for the OPD group, reliable improvement in the Hs, Hy, and Pd scores for the sample of hospitalized neurotics (PN), no significant changes for the psychotics who received no electroshock therapy (PS-NS), and reliable improvements on the Hs, D, Hy, Pd, Pt, and Sc scores for the psychotics treated with electroshock (PS-S).

B. Group Differences in the Frequency and Nature of Response Changes

Table 6 gives the mean per cent of responses changed per item for each of the groups. These values are the same as the per cent of all item-responses per group which are changed.[2] The per-

TABLE 6

Means and Standard Deviations for Per Cent of Responses Changed

Group	N	M	σ
Normals	42	14.0	10.0
OPD	25	22.0	10.8
PN	24	18.4	13.6
PS-NS	13	30.6	13.2
PS-S	20	30.6	13.4

F-values

Group	OPD	PN	PS-NS	PS-S
Normals	1.1	1.8*	1.7*	1.8*
OPD		1.6*	1.5*	1.6*
PN			1.0	1.0
PS-NS				1.0

t-values

Group	OPD	PN	PS-NS	PS-S
Normals	12.0*			
OPD				
PN			14.4*	14.0*
PS-NS				.1

* Significant at 1% level.

---

[2] Let i = no. of items, p = no. of persons in group, and c = no. of responses changed per item. Then the mean values of Table 6 are derived from $\frac{\Sigma(c/p)}{i}$, which is also equal to $\frac{\Sigma c}{pi}$.

centage of responses changed by the normals agrees well with the lower limits of comparable values reported in the literature. The percentage of responses changed in the two subgroups of normals were very similar (14.41 per cent and 15.56 per cent). It should be noted that the OPD and PN groups not only show an increase in the per cent of responses changed over that for the normals but that this per cent is practically the same for both of the therapy groups, the difference between the two values being only 3.5 per cent. Likewise, the per cent of responses changed is nearly identical for the two psychotic groups and is twice that for normals, both the shock and non-shock treated psychotics changing approximately one-third of their responses. The mean per cent of responses changed per item for the OPD group is significantly greater than that for the Normal Group. The significantly greater variances of the PN, PS-NS, and PS-S groups as compared to that of the normals prevent testing of the significance of the differences among the related means. The mean values for the PS-NS and PS-S groups are significantly greater than that for the PN group. The difference between the means of the two psychotic groups is not reliable. An hypothesis suggested by the data of Table 6 is that neurotic and psychotic patients have specific frequencies of response change to the MMPI items which are significantly different from one another, significantly greater than that for normals, and relatively independent of the specific type of therapy they are receiving.

TABLE 7

Percentage Distributions of Item Response Changes in Each Group According to Direction of Change

Group	T to F	F to T	T to ?	? to T	F to ?	? to F
Normal	37.3	33.8	6.2	5.2	9.8	7.7
OPD	42.2	43.7	2.4	4.2	2.2	5.3
PN	46.5	40.4	1.7	3.1	3.2	5.1
PS-NS	45.9	36.2	1.8	5.5	1.5	9.1
PS-S	40.6	40.2	5.4	3.2	6.1	4.5
Group		T to F and F to T		T or F to ?		? to T or F
Normal		71.1		16.0		12.9
OPD		85.9		4.6		9.5
PN		86.9		4.9		8.2
PS-NS		82.1		3.3		14.6
PS-S		80.8		11.5		7.7

Table 7 presents data on the frequency of response change in each of the groups studied in terms of the direction of the changed response. Two facts are immediately apparent from this table. First of all, there appear to be no gross differences among the groups in the directional distribution of the response changes. For each group, the percentage of responses changed from "True" to "False" is within ten points of the percentage of "False" to "True" responses. Secondly, the percentage of response changes which involve directly opposite responses, i.e., from "True to False" or "False to True," constitute, in all groups, 70 per cent or better of all changes. It may be noted that changes involving a shift from an originally uncertain attitude or introspection (as evidenced by a "Cannot Say" response) to a definitely positive or negative ("True" or "False") response constitute from 7.7 per cent to 12.9 per cent of the response changes in the five groups. Since the number of "True" and "False" responses on the first administrations of the MMPI constitute, on the average, 97 per cent of all responses, with the remaining 3 per cent of responses in the "Cannot Say" category, it is clear from Table 7 that original "True" and "False" responses contribute a smaller proportion of the total of changed responses than would be expected from the proportion they constitute of all original responses. On the other hand, the per cents of all response changes which are in the "Cannot Say" to "True" or "False" direction are from two to three times the proportions which the "Cannot Say" responses are of all original responses. From this data, it appears that the "Cannot Say" responses, which are not scored on any of the MMPI scales, are considerably less stable than the "scorable" "True" and "False" responses. This finding is in agreement with that of Lentz (13) who found the proportion of "?" answers changed on the Bernreuter inventory to be approximately twice the proportions of "Yes" and "No" answers which were changed. These findings indicate that any investigation of the differentiating potentiality of "Cannot Say" responses relative to certain classes of psychiatric patients would most likely show such responses to reflect variable rather than stable aspects of the personality. The average proportion of changed responses which are in the "Cannot Say" to "True" or "False" direction is 10.0 per cent for the four clinical groups: this value may be compared with one of 12.9 per cent for the normals. There is no indication that therapy-induced changes in personality are reflected in an increased tendency for originally uncertain attitudes to become either positive or negative.

The means and standard deviations for the number of item-responses changed per individual for each of the groups is given in Table 8. The means for the two subgroups of normals were 71.37 and 77.04 respectively, with standard deviations of 19.36 and 26.24. The differences between the means and variances of these two subgroups is not significant ($t = 2.424$; $F = 1.818$). Again, as was noted for the over-all per cent of responses changed, there is a remarkable similarity of the OPD and PN groups and of the PS-NS and PS-S groups with respect to the mean number of responses

TABLE 8

MEANS AND STANDARD DEVIATIONS FOR NUMBER OF ITEM-RESPONSE CHANGES PER INDIVIDUAL

Group	N	M	σ
Normals	42	74.4	23.6
OPD	25	108.6	33.0
PN	24	106.5	29.4
PS-NS	13	151.6	61.6
PS-S	20	151.6	37.3

F-values

Group	OPD	PN	PS-NS	PS-S
Normals	2.0	1.6	7.2*	2.6*
OPD		1.2	3.6*	1.2
PN			4.6*	1.6
PS-NS				2.8

t-values

Group	OPD	PN	PS-NS	PS-S
Normals	19.2*	18.6*		
OPD		.8		13.3*
PN				10.2*

* Significant at 1% level.

changed per individual. The mean for the Normal Group is significantly less than that of the OPD group and also significantly smaller than that of the PN group. The variances of the PS-NS and PS-S groups differ significantly from that of the normals and hence testing of the related mean differences is not applicable. The mean for the OPD group does not differ significantly from that of the PN group, and the means of the PS-NS and PS-S groups do not differ significantly. The PS-S group mean is significantly larger than that of either the OPD or the PN group.

TABLE 9

PERCENT OF ITEM-RESPONSES CHANGED PER GENERAL CATEGORY OF ITEM FOR EACH GROUP*

Category	Normals	OPD	PN	PS-NS	PS-S
I	15.3 (9)	24.0 (7.5)	21.8 (14)	31.6 (10)	38.3 (2)
II	8.0 (22)	23.4 (13)	24.3 (7)	30.8 (12.3)	30.8 (10)
III	16.0 (8)	16.7 (22)	18.9 (18)	34.3 (4)	25.0 (18.3)
IV	7.5 (23)	30.7 (1)	31.9 (2)	37.2 (1)	25.8 (16.5)
V	7.1 (24)	26.4 (3)	35.8 (1)	29.2 (18.5)	32.0 (8)
VI	10.7 (17)	19.2 (17)	22.5 (12)	26.2 (20)	27.0 (14)
VII	10.5 (19)	16.8 (20.5)	23.3 (9.5)	29.2 (18.5)	19.0 (23)
VIII	6.5 (25)	24.0 (7.5)	27.7 (3)	23.8 (21)	25.0 (18.3)
IX	11.9 (13.5)	29.6 (2)	19.2 (17)	32.3 (8.5)	21.0 (22)
X	13.5 (12)	19.4 (16)	21.9 (13)	34.0 (5)	30.3 (12)
XI	10.6 (18)	20.0 (15)	15.4 (22)	29.3 (17)	23.7 (21)
XII	22.4 (1)	22.4 (14)	27.1 (4)	36.3 (2)	39.2 (1)
XIII	11.5 (16)	19.0 (18)	14.9 (24)	23.1 (22)	25.0 (18.3)
XIV	18.3 (5)	23.8 (10)	26.0 (5)	29.8 (15)	31.3 (9)
XV	18.0 (6)	16.4 (23)	17.3 (19)	31.2 (11)	25.8 (16.5)
XVI	21.0 (2)	23.6 (11)	21.6 (15)	32.9 (7)	30.7 (11)
XVII	17.6 (7)	23.5 (12)	23.3 (9.5)	30.8 (12.3)	34.5 (5)
XVIII	11.9 (13.5)	25.3 (4)	23.0 (11)	30.8 (12.3)	34.7 (4)
XIX	18.9 (4)	24.8 (6)	23.8 (8)	34.6 (3)	34.0 (7)
XX	14.1 (11)	17.6 (19)	16.1 (20)	32.3 (8.5)	29.0 (13)
XXI	9.4 (21)	16.1 (24)	15.9 (21)	29.5 (16)	26.1 (15)
XXII	14.5 (10)	23.9 (9)	21.0 (16)	30.2 (14)	34.3 (6)
XXIII	11.6 (15)	12.0 (25)	10.1 (25)	19.8 (23)	18.6 (24)
XXIV	20.7 (3)	25.0 (5)	25.3 (6)	33.3 (6)	35.3 (3)
XXV	9.5 (20)	16.8 (20.5)	15.0 (23)	16.4 (24)	23.7 (20)

* The figures in parentheses give the ranks of the categories in the order of frequency of changed responses within each clinical group.

In an attempt to determine whether there were any group differences in the pattern of response-changes in terms of the general types of items, the percentage of item-responses changed in each of the general categories of items, as originally classified by the MMPI authors, was determined for each of the groups. These data are given in Table 9, together with the rank of each category for frequency of response change within each of the groups. The general categories plus the number of items classified in each are as follows:

I. General Health (9 items).
II. General Neurologic (19 items).
III. Cranial Nerves (11 items).
IV. Motility and Coordination (6 items).
V. Sensibility (5 items).
VI. Vasomotor, Trophic, Speech, Secretory (10 items).
VII. Cardiorespiratory (5 items).
VIII. Gastrointestinal (11 items).
IX. Genitourinary (6 items).
X. Habits (20 items).
XI. Family and Marital (29 items).
XII. Occupational (18 items).
XIII. Educational (12 items).

XIV. Sexual Attitudes (19 items).
XV. Religious Attitudes (20 items).
XVI. Political Attitudes—Law and Order (46 items).
XVII. Social Attitudes (72 items).
XVIII. Affect, Depressive (32 items).
XIX. Affect, Manic (24 items).
XX. Obsessive, Compulsive (15 items).
XXI. Delusions, Hallucinations, Illusions, Ideas of Reference (31 items).
XXII. Phobias (29 items)
XXIII. Sadistic, Masochistic (7 items).
XXIV. Morale (33 items).
XXV. Items modeled after suggestions of Hartshorne, May and Shuttleworth to indicate whether the individual is trying to place himself in an improbably acceptable or unacceptable light (15 items).

Application of the χ^2-technique revealed no significant difference in the distributions of frequencies of response changes per general category for the two subgroups of normals ($\chi^2 = 27.34$; $P > .30$).

While the most "unstable" categories in the normal subjects are XII (Occupational) and XXIV (Morale), those with greatest frequency of response change in the OPD and PN groups are IV (Motility and Coordination) and V (Sensibility). Greatest frequency of change for the PS-NS and PS groups is in XII (Occupational). There is no clear differential pattern of frequency of response change per category of item. In general, the first nine categories, those dealing with physical symptomatology of the various organ systems, Category XVIII (Affect, Depressive) XXI (Delusions, Hallucinations, Illusions, etc.) XXII (Phobias), and XXV (the "L" scale items) show a considerably greater frequency of response change in the clinical groups than characterizes the normals.

C. ITEM CHANGE IN RELATION TO SCORE CHANGE

In order to make an appraisal of the relationship between the amount of item-response change and the degree of profile change, the profiles of the three individuals having the largest number of item-response changes in each group were studied together with the profiles of the three individuals with the smallest numbers of response changes. It was at once apparent that no close relationship existed between the number of item responses changed and the degree of profile change.

This brief examination suggested that there were individual differences and probably group differences in the relative specificity of item changes with respect to profile changes which might be interpreted as reflecting therapeutic progress. It was thought advisable to devise a measure of the relationship of item-change to scale-score change and to study the distribution of this measure in each of the special groups. The measure used for this purpose was simply the ratio of the total number of item-responses changed per individual to the total amount of change, without regard to sign, in the raw scores on the clinical variables. Table 10 reports the means and standard deviations of this ratio for the normals and for each of the clinical groups.

Since only 213 of the 495 items of the MMPI included in this study are scored on any of the eight psychiatric variables covered by those items (Hs, D, Hy, Pd, Pt, Sc, and Ma), only 43 per cent of item changes would be expected to be reflected in scale score change. Thus, for every 10 item-responses changed, a minimum change of 4.3 score points would be expected on the clinical scales, if the changes in item-response were random or chance. If, on the other hand, the item-changes were in response to particular therapeutic efforts directed at the deviations measured by certain of the psychiatric scales, the ratio of item-change to score-change should be lessened, i.e., a given number of item-changes should result in a greater amount of scale-score change than is true for normals.

The findings are reported in Table 10.

The Normal Group shows a mean ratio of item-change to score-change which is just about what is to be expected by chance. The means for the clinical groups are smaller, indicating an increasing specificity of item-change, an increasingly higher proportion of scale-score items among those which are changed. The mean ratios for the PS-NS and PS-S groups are significantly smaller than that for the Normal Group. The difference between the mean ratios of the OPD and PN groups is not significant nor is that between the means of the PS-NS and PS-S groups.

TABLE 10

MEANS AND STANDARD DEVIATIONS FOR RATIO OF NUMBER OF ITEMS CHANGED TO AMOUNT OF SCORE CHANGE

Group	N	M	σ
Normals	42	4.0	2.5
OPD	25	3.3	1.0
PN	24	3.2	1.2
PS-NS	13	3.4	1.4
PS-S	20	3.4	1.7

F-values

Group	OPD	PN	PS-NS	PS-S
Normals	5.8*	3.8*	3.0	2.1
OPD		1.5	1.9	2.6
PN			1.2	1.8
PS-NS				1.4

t-values

Group	OPD	PN	PS-NS	PS-S
Normals			2.6*	3.6*
OPD		1.3	.3	.4
PN			1.1	.9
PS-NS				.0

* Significant at the 1% level.

One sees, then, that neurotic and psychotic patients undergoing various therapies not only tend to change a significantly greater number of their item-responses than do normals (Table 8), but that their response changes show a significantly greater concentration of items which are scored on the psychiatric scales of the MMPI. We have in this data positive evidence of the potential validity of the MMPI for the measurement of therapeutic effects. At the same time, the lack of close relationship between over-all amount of item change and profile change argues for caution in the interpretation of profile changes. For example, if an alcoholic takes the MMPI when not fully recovered from the general confusional and somatic effects of a severe drinking bout and repeats the inventory later with a more clear sensorium, one would expect a very large number of response changes and probably considerable profile change. No matter how extensive the profile changes might be under these conditions, they should not be interpreted as evidence of a change in personality but rather as evidence of recovery from an acute disturbance of certain perceptual-cognitive and somatic functions. Thus, if the F-score of the original profile were very high, it need not necessarily invalidate the profile in the usual sense, but rather might better be interpreted as indicating that the clinical profile reflects an acute disturbance and does not typify the individual's basic personality. If, on the other hand, this same person were tested and retested when fully sober, over an interval in which intensive psychotherapy had been attempted, an equally extensive change in his MMPI profile might be found to result from a change in response to considerably fewer items than were changed in the above example.

True *personality* changes as reflected in profile changes should mean a greater-than-chance frequency of change of those items which yield scale scores.

D. Characteristics of the Most Stable Items and the Most Variable Items in Each Group

Since the variability of the items (percentage of responses per item changed) shows a large range in each of the groups (e.g., 0-42.99 per cent in the normals), it was considered desirable to determine what differences might characterize the most stable and the most variable items for each distribution. Table 11 presents the results of an analysis of the scale representations of the two types of item in each group. Thus, it was found that the 27 items which had a 0 per cent change in the Normal Group (i.e., each individual's second response to an item was the same as his first response) yielded a total of 38 raw scale-score units.[3] Of the total amount of scale-score yielded by these items, the largest percentages are found on the F and Sc scales. A similar relationship, concentration of the most stable items on the F and Sc scales, holds in each of the four clinical groups.

No equally consistent pattern is found among the groups for the most variable items. In the normals and PS-NS groups, the Hy scale has the highest concentration of variable items. In the PN group, the Hy and Sc scales have the highest—and equal—concentrations of variable items. In the PS-S group, the D scale shows the highest concentration of variable items, with the Hy and Pd scales showing the second highest frequency. In short, the F and Sc scales are marked in all groups by a relatively high frequency of very stable items; the variable items, on the other hand, are irregular in their scale concentration from group to group. In view of the stability characterizing many of the items composing the F and Sc scales, one would have reason to expect any change in the scores on those scales to more probably reflect valid personality changes rather than to be chance variations.

Table 12 reports for each group the proportions of the most stable and most variable items for that group which are "obvious," "subtle," and not scored on

TABLE 11

PERCENTAGE DISTRIBUTIONS, BY SCALE, OF THE SCORES YIELDED BY THE MOST STABLE AND BY THE MOST VARIABLE ITEMS FOR EACH GROUP

Group	Item Type	No. of Items	% Change	Tot. Scale Score	Percent of Total Scale-Score Per Scale											
					L	F	K	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma
Normal	Stable	27	0	38	5.3	34.2		5.3	5.3	7.9	2.6	2.6	10.5	5.3	18.4	2.6
	Variable	28	35.0+	25		4.0	20.0		8.0	24.0	12.0	4.0	12.0	4.0		12.0
OPD	Stable	30	0–4	39	5.1	48.7		2.6	2.6	2.6	7.7		12.8		17.9	
	Variable	37	40.0+	31			16.1		9.7	9.7	16.1	9.7	12.9	9.7		16.1
PN	Stable	49	0–4.2	55	3.6	47.3		1.8	1.8	1.8	5.5	1.8	10.9	1.8	23.6	
	Variable	33	41.7+	45	2.3		4.4	11.1	11.1	17.8	8.9	4.4	6.7	11.1	17.8	4.4
PS-NS	Stable	28	0–7.7	31	9.7	22.6		6.5	9.7	9.7	3.2	12.9		6.5	16.0	3.2
	Variable	37	53.8+	35			14.3		8.6	17.0	14.3	2.9	14.3	8.6	5.7	14.3
PS-S	Stable	20	0–5	30	6.7	53.3					6.7		6.7	3.3	23.3	
	Variable	32	55.0+	37			5.4	8.1	21.6	16.3	16.3		5.4	10.7	8.1	8.1

---

[3] That the number of scale-score units should exceed the number of items is accounted for by the fact that certain of the items are scored on more than one scale. Lists of the most stable items, i.e., those with "0%" change, are provided in the Appendix.

tric scales, the ratio of item-change to score-change should be lessened, i.e., a given number of item-changes should result in a greater amount of scale-score change than is true for normals.

The findings are reported in Table 10.

The Normal Group shows a mean ratio of item-change to score-change which is just about what is to be expected by chance. The means for the clinical groups are smaller, indicating an increasing specificity of item-change, an increasingly higher proportion of scale-score items among those which are changed. The mean ratios for the PS-NS and PS-S groups are significantly smaller than that for the Normal Group. The difference between the mean ratios of the OPD and PN groups is not significant nor is that between the means of the PS-NS and PS-S groups.

TABLE 10

MEANS AND STANDARD DEVIATIONS FOR RATIO OF NUMBER OF ITEMS CHANGED TO AMOUNT OF SCORE CHANGE

Group	N	M	σ
Normals	42	4.0	2.5
OPD	25	3.3	1.0
PN	24	3.2	1.2
PS-NS	13	3.4	1.4
PS-S	20	3.4	1.7

F-values

Group	OPD	PN	PS-NS	PS-S
Normals	5.8*	3.8*	3.0	2.1
OPD		1.5	1.9	2.6
PN			1.2	1.8
PS-NS				1.4

t-values

Group	OPD	PN	PS-NS	PS-S
Normals			2.6*	3.6*
OPD		1.3	.3	.4
PN			1.1	.9
PS-NS				.0

* Significant at the 1% level.

One sees, then, that neurotic and psychotic patients undergoing various therapies not only tend to change a significantly greater number of their item-responses than do normals (Table 8), but that their response changes show a significantly greater concentration of items which are scored on the psychiatric scales of the MMPI. We have in this data positive evidence of the potential validity of the MMPI for the measurement of therapeutic effects. At the same time, the lack of close relationship between over-all amount of item change and profile change argues for caution in the interpretation of profile changes. For example, if an alcoholic takes the MMPI when not fully recovered from the general confusional and somatic effects of a severe drinking bout and repeats the inventory later with a more clear sensorium, one would expect a very large number of response changes and probably considerable profile change. No matter how extensive the profile changes might be under these conditions, they should not be interpreted as evidence of a change in personality but rather as evidence of recovery from an acute disturbance of certain perceptual-cognitive and somatic functions. Thus, if the F-score of the original profile were very high, it need not necessarily invalidate the profile in the usual sense, but rather might better be interpreted as indicating that the clinical profile reflects an acute disturbance and does not typify the individual's basic personality. If, on the other hand, this same person were tested and retested when fully sober, over an interval in which intensive psychotherapy had been attempted, an equally extensive change in his MMPI profile might be found to result from a change in response to considerably fewer items than were changed in the above example.

True *personality* changes as reflected in profile changes should mean a greater-than-chance frequency of change of those items which yield scale scores.

D. Characteristics of the Most Stable Items and the Most Variable Items in Each Group

Since the variability of the items (percentage of responses per item changed) shows a large range in each of the groups (e.g., 0-42.99 per cent in the normals), it was considered desirable to determine what differences might characterize the most stable and the most variable items for each distribution. Table 11 presents the results of an analysis of the scale representations of the two types of item in each group. Thus, it was found that the 27 items which had a 0 per cent change in the Normal Group (i.e., each individual's second response to an item was the same as his first response) yielded a total of 38 raw scale-score units.[3] Of the total amount of scale-score yielded by these items, the largest percentages are found on the F and Sc scales. A similar relationship, concentration of the most stable items on the F and Sc scales, holds in each of the four clinical groups.

No equally consistent pattern is found among the groups for the most variable items. In the normals and PS-NS groups, the Hy scale has the highest concentration of variable items. In the PN group, the Hy and Sc scales have the highest—and equal—concentrations of variable items. In the PS-S group, the D scale shows the highest concentration of variable items, with the Hy and Pd scales showing the second highest frequency. In short, the F and Sc scales are marked in all groups by a relatively high frequency of very stable items; the variable items, on the other hand, are irregular in their scale concentration from group to group. In view of the stability characterizing many of the items composing the F and Sc scales, one would have reason to expect any change in the scores on those scales to more probably reflect valid personality changes rather than to be chance variations.

Table 12 reports for each group the proportions of the most stable and most variable items for that group which are "obvious," "subtle," and not scored on

TABLE 11

Percentage Distributions, by Scale, of the Scores Yielded by the Most Stable and by the Most Variable Items for Each Group

Group	Item Type	No. of Items	% Change	Tot. Scale Score	Percent of Total Scale-Score Per Scale											
					L	F	K	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma
Normal	Stable	27	0	38	5.3	34.2		5.3	5.3	7.9	2.6	2.6	10.5	5.3	18.4	2.6
	Variable	28	35.0+	25		4.0	20.0		8.0	24.0	12.0	4.0	12.0	4.0		12.0
OPD	Stable	30	0–4	39	5.1	48.7		2.6	2.6	2.6	7.7		12.8		17.9	
	Variable	37	40.0+	31			16.1		9.7	9.7	16.1	9.7	12.9	9.7		16.1
PN	Stable	49	0–4.2	55	3.6	47.3		1.8	1.8	1.8	5.5	1.8	10.9	1.8	23.6	
	Variable	33	41.7+	45	2.3		4.4	11.1	11.1	17.8	8.9	4.4	6.7	11.1	17.8	4.4
PS-NS	Stable	28	0–7.7	31	9.7	22.6		6.5	9.7	9.7	3.2	12.9		6.5	16.0	3.2
	Variable	37	53.8+	35			14.3		8.6	17.0	14.3	2.9	14.3	8.6	5.7	14.3
PS-S	Stable	20	0–5	30	6.7	53.3					6.7		6.7	3.3	23.3	
	Variable	32	55.0+	37			5.4	8.1	21.6	16.3	16.3		5.4	10.7	8.1	8.1

---

[3] That the number of scale-score units should exceed the number of items is accounted for by the fact that certain of the items are scored on more than one scale. Lists of the most stable items, i.e., those with "0%" change, are provided in the Appendix.

any scale. The "obvious" and "subtle" items are composed of all items scored on the D, Hy, Pd, Pa, and Ma scales of the MMPI and have been keyed by Wiener and Harmon (18) according to the relative ease or difficulty with which the nature of the socially desirable response may be detected. It will be noted that there is a consistent tendency for the stable items to consist of higher proportions of obvious rather than subtle items, with a mean difference of 15.3 per cent between the proportions of these two types of item among the stable items of the five groups. No such consistent tendency appears in the obvious-subtle distributions of the variable items. In comparing the stable and variable items, no consistent group-to-group relationship appears with respect to the frequency of obvious items. However, in all groups, the variable items show higher proportions of subtle items than are found in the stable items, and for the Normal and PS-NS groups the differences between the proportions of subtle items in the stable and variable items respectively are significant *($P < .01$)*. Furthermore, there is a consistent tendency for the variable items to have a higher proportion of items which are not scored on any scale than is true for the stable items, and the differences between the proportions of non-scored items in the stable and variable items is significant for the Normal and OPD groups. In short, the variable items, as distinguished from the stable items, are characterized by a tendency to be either of a type in which the direction of the socially desirable response is not readily perceptible or of a type which has not yielded normal-criterion critical ratios sufficiently large to result in their selection for scoring on any of the existing MMPI scales. That items having a high frequency of changed responses upon repetition with a given group should show these characteristics is to be expected both from the psychology of that behavior which is variable and from the nature of item-selection. Thus, there are areas of behavior the "subtlety" of which is expressed by the absence of any clear-cut, readily recognized pattern of response in the group which may serve as a constant standard against which the individual can evaluate or regulate his own conduct. In the absence of such social norms, the individual is required to "project" norms which vary as a function of recent experience and memory of last response in a similar situation. The element of uniqueness in all experience plus the fallibility of memory makes directly for variability of response in the individual.

TABLE 12

THE PROPORTIONS OF THE MOST STABLE AND MOST VARIABLE ITEMS FROM EACH GROUP WHICH ARE "OBVIOUS," "SUBTLE," AND NOT SCORED ON ANY SCALE

Group	The Most Stable Items				The Most Variable Items			
	No. of Items	Percent Obvious	Percent Subtle	% Not Scored	No. of Items	Percent Obvious	Percent Subtle	% Not Scored
Normal	27	25.9	0.0	14.8	28	10.7	32.1	47.9
OPD	30	23.3	6.7	13.3	37	27.0	16.2	48.6
PN	49	12.2	6.1	20.4	33	27.3	21.2	21.2
PS-NS	28	21.4	3.6	28.5	37	21.6	32.4	30.0
PS-S	20	10.0	0.0	20.0	32	40.6	15.6	40.6

This variability expresses itself in the group by an absence of a given response in sufficient frequency to characterize that group in statistical comparison with another group.

As an aside, it might be proposed that the ideal questionnaire should be composed solely of items of the subtle variety which require the subject to "project" his own norms, thus gaining for the personality inventory the highly regarded virtues of the projective techniques. There are at least two answers to this proposal. In the first place, the variability in response to such items, as has been indicated in this study, would suggest that they might seriously detract from the effort to establish reliable relationships between the test measures and given criteria. Secondly, recognizing that there are attitudes peculiar to the process of test-taking which are different from the particular attributes we are interested in measuring and which have certain distorting effects on the latter, there would seem to be some inherently scientific cogency to permitting such distorting effects to operate in such a way that they may be measured, rather than naively assuming that they have been circumvented by a clever ruse.

Further study of the stable and variable items was made in an effort to discover any distinguishing characteristics that might not have appeared in the above statistical analyses. Other studies (10) have indicated that items involving an estimate of frequency by the subject (e.g., "much of the time," "seldom," etc.) have higher rates of change than items not requiring such estimates. An analysis was made of the occurrence of these two types of item among the stable and variable items of each of the groups in the present investigations, with the results reported in Table 13. A consistent tendency was found for the stable items of each group to have a higher proportion of items not involving an estimate of frequency than was true for the variable items. For the stable items from all groups, the mean ratio of the number of items not involving an estimate of frequency to the number of items involving

TABLE 13

THE PROPORTIONS OF ITEMS INVOLVING FREQUENCY ESTIMATES AND ITEMS NOT INVOLVING SUCH ESTIMATES IN THE STABLE AND VARIABLE ITEMS OF EACH GROUP

Group	Type of Item	No. of Items	% With Freq. Estimate	% Without Freq. Estimate	Ratio of No. of Non-Freq. to Freq. Items
Normal	Stable	27	23.1	76.9	3.33
	Variable	28	33.3	66.7	2.00
OPD	Stable	30	16.7	82.3	5.00
	Variable	37	40.5	59.5	1.47
PN	Stable	49	14.3	85.7	6.00
	Variable	33	38.2	61.8	1.62
PS-NS	Stable	28	31.0	69.0	2.33
	Variable	37	37.8	62.2	1.64
PS-S	Stable	20	10.0	90.0	9.00
	Variable	32	28.1	71.9	2.55

such an estimate was 5.13; for the variable items, the comparable mean ratio was 1.86. It appears from this that there is a distinct tendency for those items in which the subject must make some estimate of frequency of occurrence of a behavior in himself or others (e.g., "I feel weak all over much of the time," or "I believe I am no more nervous than most others") to be more variable in the responses they elicit from the same individuals than are the items not requiring such estimates.

In a study of the relationship of variability of items and the degree to which they were stated in present or past tense, no consistent group-to-group tendency was found. In three of the five groups (Normal, PN, and PS-S), the variable items yielded a higher ratio of "number of items in the present tense" to "number of items in the past tense" than did the stable items. In the other two groups (OPD and PS-NS), this relationship was reversed, the stable items yielding a higher ratio of present-tense to past-tense items. The mean of these ratios in the five groups was 4.10 for the stable items and 8.10 for the variable items. As might be expected, those items involving a memory of past occurrences appear more frequently among those items to which responses are quite stable, while those items involving present situations, moods, and pressures appear more frequently among those items to which responses are changed. However, the variability in the relative concentration of present-tense and past-tense items among the stable and variable items from group to group suggests that the relationship between the "time" factor of an item and its stability or variability is not a simple one but involves other elements.

Finally, a study was made to determine the relative stability of responses to the so-called "double-negative" items. These are items which are stated in the negative and are scored if answered in the negative, that is sorted "False" (e.g., "I do not tire quickly"). The stable and variable items from each of the five groups of subjects were examined to determine the number of double-negative items they included. This survey of the 154 stable and 157 variable items yielded a total of 33 items of the double-negative variety. Of these 33 items, 10 (or 30.3 per cent) were stable and the remainder, 69.7 per cent, were variable. In four of the five groups, the ratio of the number of double-negative items among the variable items to the number among the stable items was three to one. In the remaining group (PS-NS), the ratio was one to one.

In summary, the above analyses of the most stable and most variable items for each of the five groups of subjects have pointed to certain fairly consistent differentiating characteristics. The stable items tend to be concentrated on the F and Sc scales of the MMPI, to be obvious in nature, to involve no estimate of the frequency of occurrence of a given behavior either in the subject or in other people, and to be more frequently stated in past rather than present tense. The variable items, on the other hand, tend to show no clear concentrations on any of the MMPI scales, to be subtle rather than obvious, to involve estimates of frequency of given behavior, and to be stated in the present rather than the past tense.

E. INVESTIGATION OF A THERAPY-SUSCEPTIBLE SCALE

It was suggested earlier in this research that one possible direction to be taken by a study of item-response changes would be a determination of those items which have a significantly greater percentage of changed responses in persons responding favorably to therapy than in normals. Theoretically, such a scale of items would have both a prognostic and an evaluative function. For example, the greater the clinical score made by the individual on those items having a specific tendency to change with electroshock therapy, the greater the probability that his post-therapy responses will be in the non-psychiatric direction, i.e., the greater the probability that he will improve. It follows directly from this that the amount of the individual's change in score on the therapy-susceptible items may be taken as an index of the validity of the overall profile change as an indication of improvement. Profile changes in the favorable direction not accompanied by adequate changes on the items specifically responsive to the therapy concerned would not be misinterpreted as evidence of improvement.

It might be argued that such proposed scales of therapy-susceptible items should be composed solely of those items for which the frequency of change of an originally significant response in a given therapy group was reliably greater than the same change in normals. Such a selection procedure would miss those equally variable items which happened on the first test to have been answered in the nonsignificant direction. Consequently, the following procedure was used to arrive at a list of "therapy-susceptible" items for each of the treatment groups.

The total number of response-changes was determined for each item for each of the "improved" clinical groups (PN, PS-NS, PS-S) and for the combined normals.[4] These frequencies were converted to percentages. By the use of appropriate nomographs (19) the reliability of the difference between the percentage of changed responses for each item in the Normal Group and in a given therapy group was determined. Only those differences were considered reliable which yielded a critical ratio of 3.00 ($P < .003$). In the cases of percentage differences which were borderline with respect to the nomographs, the following formula was applied:

$$t=\frac{p_1-p_2}{\sqrt{\frac{N_1p_1+N_2p_2}{N_1+N_2}\left(1-\frac{N_1p_1+N_2p_2}{N_1+N_2}\right)\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}}$$

By this procedure, three scales of "therapy-susceptible" items were derived. In terms of the arguments given above, the items of these scales may be considered to constitute a potential measure of prognosis and the scales will be referred to hereafter as the Px scales.

Since the hospitalized neurotics appeared to be the group for which independent validating data could be most readily obtained, it was decided to study the prognostic effectiveness of the scale of items derived for this group.[5] A key based on these items was applied to both MMPI records of the 24 patients in the PN group. Basic statistics for the Px scores in this group are given in Table

---

[4] It has been reported previously that the χ^2 technique showed no significant differences between the two normal groups in the manner in which item changes were distributed over the twenty-five general categories of items.

A study of the significance of the difference of frequency of response-change per item in the two normal groups revealed a total of only 15 items which yielded differences significant at the 5 per cent level. Of these 15 items, only 5 (1 per cent of the total number of items) yielded differences significant at the 1 per cent level. In view of these findings, it was considered permissible to combine data from the two normal groups for the item study.

[5] The twenty-four items comprising this Px scale, as numbered in the MMPI manual of directions, 1943 version, are as follows: A2, A4, A13, A16, A46, A48, B1, B9, B11, B12, B14, B15, B17, B28, C43, F2, F18, F49, F55, G1, G20, G21, H46, I27.

14. The mean Px score on the second MMPI is significantly lower than that obtained in the first administration. While this is evidence of the potential usefulness of the Px scores for detecting therapeutic changes on a group basis, the extreme overlap of scores before and after therapy makes it practically useless for individual prediction.

TABLE 14

MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE TESTS FOR THE Px SCORE IN THE HOSPITALIZED NEUROTICS (PN) ($N=24$)

	M	σ	Range
1st MMPI....	11.9	3.5	4–20
2nd MMPI....	8.1	4.8	1–21
$t=3.5$ $P<.01$		$F=2.1$ $P<.05>.02$	

In order to study the distribution of the Px score in independent samples, the hospital files were searched for cases of diagnosed female psychoneurotics who were discharged "unimproved." Only thirty such cases were found and of this number, only nineteen had MMPI records. Twenty-nine cases of female psychoneurotics discharged "improved" and having post-admission MMPI records were also chosen randomly from the files. Table 15 gives the basic statistics of the Px score for these two groups.

The difference between the mean Px scores of the two groups is not significant.

In addition to the information on "condition on discharge" for the two test groups, the diagnostic summary of each patient carried a statement relative to prognosis. These statements were evaluated in terms of "good," "fair," or "poor" prognosis. The χ^2 technique revealed no relationship between the Px score at the time of admission and the prognosis on discharge in either the "improved" or "unimproved" groups ($\chi^2 = .939$ and $.254$, respectively).

In order to determine whether any better differentiation would result from the use for the Px scale of only those items for which the frequency of change of an originally significant response was reliably greater in the PN group than in the Normal Group, a key was prepared for the 11 items from the 24 of the first Px scale which satisfied this relationship. These items were: A-2, A-13, A-46, B-9, B-11, B-14, B-17, B-28, F-18, and I-27. The results of the application of this scale to the two test groups of hospitalized neurotics revealed no improvement in differentiation over that of the original Px score ($t = .446$; $F = 1.076$).

It would appear from the above data that the general approach to the selection of therapy-susceptible items which has been followed in this study does not result in a valid set of items having any prognostic power to differentiate patients who will improve under a given regimen from those who will not. This conclusion necessarily assumes that the criterion used to test the validity of such items was in itself a valid criterion.

TABLE 15

MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE TESTS FOR THE Px SCORE IN SAMPLES OF IMPROVED AND UNIMPROVED PSYCHONEUROTICS

	M	σ	Range
Improved (N = 29)..	10.6	4.6	3–20
Unimproved (N = 19)	8.6	5.1	0–18
$t=1.4$ $P<.20>.10$	$F=1.2$ $P>.05$		

In the nine years that the Psychopathic Unit of the University Hospitals has been in operation, a total of 470 female admissions has been diagnosed as psychoneuroses. Of this total, only 6.4 per cent were discharged "unimproved." A recovery or improvement rate of better than 90 per cent is so much greater than the median value reported in the literature that one would necessarily suspect a biased selection of patients, a therapeutic program of particular composition and intensity, an inaccurate evaluation of post-therapy status, or an admixture of these factors.

Careful evaluation of the histories of the neurotic patients in this study and of the larger hospital sample of neurotics from which they were drawn does not support the hypothesis of biased selection of case material. Review of the procedures used in the treatment of patients by the Psychopathic Unit and analysis of the hospital regimen, aside from particular therapeutic techniques, does not support the hypothesis that the greater improvement rate in the patients of this study as contrasted with those reported in the literature is to be accounted for by differences in treatment. There remains the probability that the high frequency of improved patients chiefly reflects errors of omission in the clinical evaluation of the status of patients and the amount of their change at the time of discharge. Support to this point is given by a readmission rate of 10.5 per cent for the female neurotics.

A more objective determination of the post-therapy status of patients than is usually available is greatly to be desired. Such determination ultimately must reside in the clinical judgment of the psychiatrist. To the extent that the psychiatrist can be encouraged to orient his post-therapy evaluation of a patient to a relationship with the factors underlying his assay of pre-therapy status, he will provide a valid clinical criterion. With such a criterion available, it will be possible for research techniques to establish independent prognostic and evaluative measures which will not only have value in the consideration of the individual patient but will provide even more important objective checks on the efficacy of specified therapeutic efforts.

CHAPTER IV

SUMMARY AND CONCLUSIONS

THIS STUDY has had three goals: to determine the nature and significance of changes in the MMPI profiles of groups of psychiatric patients receiving certain therapies, to study the extent and nature of item-response changes in these clinical groups and to contrast them with those found in normals, and, finally, to attempt the selection and testing of a group of therapy-susceptible items.

A check of the approximately 7000 records in the MMPI research files at the University of Minnesota yielded a total of 124 cases for which two profiles were available and which met the other criteria for this study. These cases were distributed as follows: normals, 42; out-patients, 25; hospitalized neurotics, 24; hospitalized psychotics not treated with any type of shock therapy, 13; hospitalized psychotics given electroshock therapy, 20. All in-patients included in the sample had been discharged "improved" or "recovered" and the two MMPI profiles had been obtained within the limits of a single hospitalization. Since it was not possible to find a sufficient number of male cases, the samples were composed entirely of females.

The sample of psychotic patients treated with electroshock, though a very mixed group, was composed largely of depressions and involutional melancholias. The average number of grand mal treatments received by these patients was between six and seven, and the interval between their two MMPI's was slightly less than one month. This group showed a significant post-shock reduction of their mean scores on all of the MMPI clinical scales with the exception of Masculinity-Femininity and Hypomania. Those means which were highest before therapy in this group were on the Depression, Psychopathic Deviate, Psychasthenia and Schizophrenia scales, and it was on these scales that greatest change occurred.

The sample of hospitalized neurotics had a mean initial profile practically identical with that of the out-patient sample. However, after a mean interval of approximately one month, the same as the median interval for the out-patient group, the profile for the hospitalized neurotics showed a significant decrease of the Hypochondriasis, Hysteria and Psychopathic Deviate scores. It must be remembered that all patients in the three hospitalized groups were discharged as "improved" or "recovered." It may be concluded that improvement in the personality pattern of neurotics under the general regimen of the psychiatric ward is shown in a significant reduction of hypochondriacal and hysterical symptomatology as measured by the MMPI. Apparently, this form of therapy does not suffice to reduce any marked degree of depression as reflected in the Multiphasic "D" score. The ineffectiveness of hospitalization alone to cure depressive states in a neurotic syndrome can only be judged with respect to the interval of intramural therapy represented in our sample, namely, approximately one month.

Within the limits of our small sample of psychiatric out-patients and the nature and number of psychiatric interviews intervening between the collection of the two MMPI records, it would not appear that out-patient therapy of the type afforded by the University Clinic suffices to affect a significant change in the mean scores of such patients on any of the clinical scales. The out-patient sample consisted almost wholly of psychoneurotics of mixed type who had an average of five clinic visits between their tests. No clinical evaluation of the nature of their response to therapy was available. The data of this study would indicate that such patients require more intensive therapy than that afforded by the out-patient clinic or a greater number of therapy interviews (or both) before they will show a favorable response. This conclusion is presented with certain reservations. It may be argued that improvement in the out-patient group occurred in areas of attitude, self-evaluation, and personal relationship not sampled by the MMPI. It may further be argued that the validity of the MMPI as a measure of therapeutically derived changes in personality has not been sufficiently established to warrant its application as the sole criterion in evaluating such changes. However, in the absence of other criteria and with consideration for the validity of the MMPI as a picture of the patient's pre-therapy status, the data of the present study do not yield evidence of improvement in the psychiatric status of the out-patients.

With respect to the proportion of item-responses changed, it was found that all of the clinical groups changed a larger percentage of their responses than was true for the normals, who changed only 14 per cent of their responses. Both groups of neurotics changed approximately 20 per cent of their responses, while the two psychotic groups gave a 30 per cent response change. Aside from the fact that the clinical groups show a greater variability than the normals in their responses, it is of considerable interest that both groups of neurotics show essentially the same rate of change in spite of marked differences in treatment, and that both psychotic groups likewise show a nearly identical mean frequency of response change in spite of the marked differences in the therapy they received. These findings suggest a characteristic rate of variability for normals, for neurotics, and for psychotics which is relatively independent of environmental stresses, although the items on which the variability is based are different. On a stability-of-response continuum, the neurotics would fall between the normals and the psychotics, somewhat closer to the normals. Further investigation of this finding might yield a useful technique for solving borderline neurotic-psychotic differentiations.

It has been shown that for normals and for groups of neurotics and psychotics, "Cannot Say" responses have a considerably higher rate of change than is found for either "True" or "False" responses. The proportion of all changed responses contributed by changes in original "Cannot Say" responses is from two to three times greater than the proportion which "Cannot Say" responses are of all original responses. The greater variability of the "Cannot Say" responses suggests that such responses may occur on items which pertain to components of personality which are "validly" variable rather than stable.

By selecting cases from the extremes of the distributions of number of items changed per individual, it was demonstrated than no simple relationship holds between the total number of item-responses changed and the amount of scale-score change effected. Marked changes in profile may be accomplished by individuals who have changed relatively few items, whereas other individuals may change a considerably greater number of items and yet show very little profile change.

This finding led to a study of the ratio of the number of items changed to the total amount of change in scale-score for the members of each of the groups. As a result of this investigation, it was discovered that the clinical groups not only change a significantly greater number of items than do normals, but that the items they change include a significantly greater concentration of scored items, i.e., items contributing to a score on at least one scale of the MMPI. This finding is considered evidence of the sensitivity of the MMPI to therapy-induced changes in personality.

While the most stable items for each of the groups, normal and clinical alike, tended to be concentrated in the F and Schizophrenia scales, no similar pattern of concentration on any scale or scales characterized the most variable items.

With respect to the "obvious" or "subtle" qualities of the items, stable items for each group were much more often obvious than subtle. No similar relationship was found for the most variable items, but these items were much more frequently not scored on any of the MMPI scales than were the stable items. The tendency for the most variable items to be more commonly not scored on any of the scales probably reflects the fact that when item selection is on the basis of group differentiation at a given time, items which pertain to the more subtle and fluctuating aspects of personality, and hence on which there would be no marked agreement by a given group, are missed.

The attempt to derive a scale of items which would both predict the response of a given type of patient to a given therapy and measure the amount of personality change to be attributed to that therapy was not successful. By selecting those items with a frequency of response-change reliably greater in an improved patient-therapy group than in normals and by keying such items for the psychiatrically significant response, it was believed that the resultant scale would yield scores with a positive relationship to prognosis, high scores being associated with good prognosis and low scores with poor. The scale derived by this approach for the hospitalized neurotics was applied to both sets of MMPI records for this group. In accordance with the assumptions about the therapy-measuring potential of such a scale, the mean score of the post-therapy records for the hospitalized neurotics was significantly lower than that of the pre-therapy records, although there was a marked degree of overlap in the two distributions of the score, over a fifth of the post-therapy scores being above the mean of the pre-therapy distribution.

Application of the therapy-susceptible scale for hospitalized neurotics to independent samples of patients discharged "improved" and "unimproved" respectively yielded no differentiation. It was pointed out in this connection that the extreme rarity of hospitalized neurotics who were discharged "unimproved" would give reason to question the validity of the "improved-unimproved" criterion. Within both the "improved" and "unimproved" groups no relationship was found between the score on the therapy-susceptible scale at time of admission and the prognosis at time of discharge.

Several lines of research are suggested by the results of this study. It is obvious, of course, that the analyses reported here should be repeated for data on males. With respect to the detection of nonspecific profile changes and their discrimination from therapy-reflecting changes, further study of the characteristics of profiles for people with high and low ratios of number of items changed to amount of score change should be made. The frequency of item-response changes in other samples of neurotics and psychotics should be determined as a check on the hypothesis that there is a characteristic rate of variability in these groups which might be used as a differential sign.

The problem of developing a prognostic scale for the MMPI would appear to be seriously hampered by the lack of a reliable clinical criterion of patients' responses to therapy. Until a technique may be developed whereby the "condition on discharge" will carry more exact meaning than it appears to currently, scales to predict response to therapy will be very difficult to validate.

APPENDIX

The Most Stable Items in Each Group[1]

A. THE TWENTY-SEVEN ITEMS WITH "0%" CHANGE IN THE COMBINED NORMALS ($N = 42$):

A11, A19, A40, A46, B5, B32, B46, C13, C14, C36, C49, D29, E37, G11, G35, H4, H14, H15, H20, H21, H22, H25, H30, H48, I7, J50, J53.

B. THE TWELVE ITEMS WITH "0%" CHANGE IN THE OUT-PATIENT GROUP ($N = 25$):

A51, D43, E21, E22, G44, G53, H14, H15, H20, H27, J45, J48.

C. THE FIFTEEN ITEMS WITH "0%" CHANGE IN THE HOSPITALIZED NEUROTICS ($N = 24$):

C9, C14, C46, C52, D21, D25, D43, E22, E25, H8, H15, H20, H22, H24, H41.

D. THE SEVEN ITEMS WITH "0%" CHANGE IN THE "NON-SHOCK" PSYCHOTICS ($N = 13$):

E15, E20, E49, H38, I9, J48, J51, J53.

E. THE SIX ITEMS WITH "0%" CHANGE IN THE "SHOCKED" PSYCHOTICS ($N = 20$):

C11, C14, C52, D28, H24, I9.

---

[1] The items are numbered as in the MMPI manual of directions, 1943 version.

BIBLIOGRAPHY

1. Appel, J. S., Beebe, G. W., and Hilger, D. W. Comparative incidence of neuropsychiatric casualties in World War I and World War II. *Amer. J. Psychiat.*, 1946, **103**, 196-199.
2. Benton, A. L., and Probst, K. A. A comparison of psychiatric ratings with MMPI scores. *J. abn. soc. Psychol.*, 1946, **41**, 75-78.
3. Carmichael, H. T., and Masserman, J. H. Results of treatment in a psychiatric outpatient department—a follow-up study of 166 cases. *J. Amer. med. Assoc.*, 1939, **113**, 2292-2298.
4. Darley, J. G. Changes in measured attitudes and adjustments. *J. soc. Psychol.*, 1938, **9**, 189-199.
5. Felix, R. H. The National Mental Health Act. *Ment. Hyg.*, 1947, **31**, 363.
6. Gough, H. G. Diagnostic patterns on the MMPI. *J. clin. Psychol.*, 1946, **2**, 23-27.
7. Hales, W. M., and Simon, W. Minnesota Multiphasic Personality patterns before and after insulin shock therapy, *Amer. J. Psychiat.*, 1948, **105**, 254-258.
8. Hathaway, S. R., and McKinley, J. C. A multiphasic personality schedule: I. Construction of the schedule. *J. Psychol.*, 1940, **10**, 249-254.
9. Hathaway, S. R., and McKinley, J. C. *Manual for the Minnesota Multiphasic Personality Inventory.* New York: Psychol. Corp., 1945.
10. Hertzman, M., and Gould, R. The functional significance of changed responses in a psychoneurotic inventory. *J. abn. soc. Psychol.*, 1939, **34**, 336-350.
11. Kelly, T. L. *Statistical methods.* New York: Macmillan, 1923.
12. Kubie, L. W. Elements in the medical curriculum which are essential in the training for psychotherapy. *Transactions of the First Conference on Training in Clinical Psychology,* March 27-28, 1947. Josiah Macy, Jr., Foundation.
13. Lentz, T. F. Reliability of opinionnaire technique studied intensively by the retest method. *J. soc. Psychol.*, 1934, **5**, 338-364.
14. McCann, W. H. Responses of psychotic patients to the Thurstone Personality Schedule before and after metrazol treatment. *Ind. Acad. Sci.*, 1938, **48**, 187-189.
15. Modlin, H. C. A study of the MMPI in clinical practice. *Amer. J. Psychiat.*, 1947, **103**, 758-769.
16. Pacella, B., Piotrowski, Z., and Lewis, N. D. C. The effects of electroconvulsive therapy on certain personality traits. *Amer. J. Psychiat.*, 1947, **104**, 83-91.
17. Walker, H. N. Concerning the standard error of a difference. *J. educ. Psychol.*, 1929, **20**, 53-60.
18. Wiener, D. N., and Harmon, L. R. I. Subtle and obvious keys for the Minnesota Multiphasic Personality Inventory; their development. *Office Memo #2 to Chiefs of Guidance Centers.* Veterans Administration. Vocational Rehab. & Educ. Div., Minneapolis, Minn. August, 1946.
19. Zubin, J. Nomographs for determining the significance of the difference between the frequencies of events in two contrasting series of groups. *J. Amer. stat. Assoc.*, 1939, **34**, 539-544.

VOLUME 64
NUMBER 6

WHOLE NO. 312
1950

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

A Scale for Measuring Teacher-Pupil Attitudes and Teacher-Pupil Rapport

By
CARROLL H. LEEDS
Furman University
Greenville, S.C.

Accepted for publication, October 8, 1949

Price $1.00

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION
1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

COPYRIGHT, 1950, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION

TABLE OF CONTENTS

I. THE PROBLEM AND ITS SETTING 1

II. THE CONSTRUCTION, ADMINISTRATION, AND SCORING OF THE TEACHER-PUPIL INVENTORY 3

- A. CONSTRUCTION OF THE ORIGINAL INVENTORY 3
  - 1. Nature of the Measuring Instrument 3
  - 2. Construction of the Items 3
  - 3. Classification of the Items 3
  - 4. Two Forms of the Inventory 4
- B. ADMINISTRATION OF THE ORIGINAL INVENTORY 5
  - 1. Empirical Basis for Standardization 5
  - 2. Nature of the Teacher Population 5
  - 3. Personal and Descriptive Data 6
  - 4. Explanation and Directions for Completing the Inventory 7
  - 5. Distribution of the Inventories 7
- C. COMPARISON OF RESPONSES AND SELECTION OF ITEMS 8
  - 1. Comparison of Responses to the Items 8
  - 2. Selection of Items for the Final Inventory 10
- D. THE ORIGINAL SCORING SYSTEM FOR THE INVENTORY 11
  - 1. Procedure Used in Calculation of Weights for Item Responses.. 11
- E. THE SIMPLIFIED SCORING SYSTEM FOR THE INVENTORY 11
- F. COMPARISON OF INVENTORY SCORES OF THE SUPERIOR AND INFERIOR GROUPS OF TEACHERS 11
  - 1. Distribution of Inventory Scores for the Superior and Inferior Groups of Teachers 12

III. THE VALIDITY AND RELIABILITY OF THE TEACHER-PUPIL INVENTORY 13

- A. DETERMINING THE VALIDITY OF THE INVENTORY 13
  - 1. Plan of Procedure 13
  - 2. Administration of the Revised Inventory 13
    - a. The nature of the teacher population 13
    - b. Mechanics of administration 14

3. Obtaining the Ratings of the Teachers 14
    - a. Ratings by principal 14
    - b. Classroom observation 15
    - c. Attitudes of pupils 16
    - d. The reliability of the ratings 17
        - (1) Use of split-half technique 18
        - (2) Reliability of the scale used by the principal 18
        - (3) Reliability of the scale used by the writer 18
        - (4) Reliability of the questionnaire used by the pupils 18

4. Comparisons Among Ratings and Inventory Scores 19
    - a. The distribution of Inventory scores for the unselected group of teachers 19
    - b. Correlations between Inventory scores and rating scale scores 20

5. An Additional Check on the Validity of the Items 21

B. DETERMINING THE RELIABILITY OF THE INVENTORY 22

IV. SUMMARY AND CONCLUSIONS 23

BIBLIOGRAPHY 24

A SCALE FOR MEASURING TEACHER-PUPIL ATTITUDES AND TEACHER-PUPIL RAPPORT[1]

I. THE PROBLEM AND ITS SETTING

In view of the vast amount of educational research carried on in recent years, it is rather surprising that one very important problem has almost escaped the critical attention of investigators. This is the problem of relationship between teacher and pupil, or the more specific phase of this relationship as manifested in the personal interaction of teacher and pupil in the classroom. There are far too many classrooms today in which the social atmosphere is little above the level of barbarism with reference to teacher-pupil interaction. Too often, pupils are treated with little respect, discipline problems rob the teacher of energy needed elsewhere, and classrooms are governed by individuals who are seriously maladjusted in personality. Although recent years undoubtedly have witnessed a growing interest in this problem, research workers have contributed little toward its solution, presumably because of the difficulty a scientific approach into the field would present. Modern educators and school administrators are, in most cases, fully cognizant of the problem, but a general feeling of helplessness with reference to its solution has prevented an energetic effort to remedy the undesirable situation. Granted the importance of the problem, it would seem that an instrument designed to measure certain aspects of teacher-pupil interaction would prove of great value in the hands of educators and school administrators.

The purpose of this study is to explore the possibility of constructing a measuring instrument that will aid in the differentiation of teachers who have or will have rapport with pupils, from those who do not or will not have such rapport. It is postulated, of course, that rapport between teacher and pupil constitutes one of the many factors essential to teaching success. Although only one of many such factors, it is assumed further that it is one of the most important. Rapport between teacher and pupil refers to a state of harmonious relationship characterized by mutual affection, sympathy, understanding, and cooperative behavior. Definitely involved is the teacher's personality which, on assumption, either facilitates or checks the establishing of satisfactory rapport with pupils. Numerous studies have indicated the importance of personality to success in teaching. However, the present study is not directly concerned with prognosis of success in teaching, and the personality factor, affecting teacher-pupil rapport as it does, is assumed merely as *one very important* element in both the vocational life of the teacher and the learning career of the pupil.

As attitudes and feelings probably represent the core of human relationships in all of their manifestations and as the area covered by teacher-pupil relationships is quite large, consideration, in general, will be restricted to these relations *as exemplified in the attitudes and feelings each group holds toward the other.* It will be assumed that a teacher's attitude toward pupils and toward children in general is an index to the rapport he has or will have with them.

---

[1] The major part of a Ph.D. investigation conducted at the University of Minnesota under the direction of Dr. Walter W. Cook, Professor of Education. To Dr. W. W. Cook and Dr. T. W. McConnell the writer gratefully acknowledges his indebtedness for a challenging problem and the inspiration and helpful ideas so essential to an effective program of attack.

Rapport between two people involves relationship in two directions. Not only, then, must we consider the teacher's attitude toward pupils, but also how the pupils feel toward the teacher whose attitude has been determined. In the present study the possibility is explored of validating the instrument for the measurement of a teacher's attitude by determining the attitude of the pupils toward that teacher. Thus, the reaction of the pupils toward each particular teacher in the study represents one of the criteria against which the present instrument is validated. Presumably, a relationship between teacher and pupil in which the teacher holds an unsatisfactory attitude would also be characterized by antagonism or resentment on the part of the pupil. At least, this represents a possibility which this study attempts to explore.

II. THE CONSTRUCTION, ADMINISTRATION, AND SCORING OF THE TEACHER-PUPIL INVENTORY

A. Construction of the Original Inventory

1. *Nature of the Measuring Instrument*

The Thurstone technique of scale construction (20) was rejected principally because its rigid demand for linearity along a degree-of-favorableness continuum would seem to present an undesirable restriction and limitation. Furthermore, the use of judges or experts (as in the Thurstone technique) was not contemplated in the present study.

Although the Likert technique (11) was not adopted in its entirety, the type of instrument finally decided upon is similar in some respects to that of Likert's. The multiple-choice mode of response to opinion statements (each item thus approximating a scale in itself) was employed. The assigning of numerical values, for scoring purposes, to each of the several responses per item also represents a point of similarity, although the method of weighting these responses differs considerably from that used in the Likert technique.

The instrument was to be known as the *Teacher-Pupil Inventory*.

2. *Construction of the Items*

Opinion statements were chosen in preference to concrete situations with multiple-choice reactions. It was felt that a relatively large number of opinion statements, touching upon a variety of topics, would tap more adequately a teacher's attitude toward pupils than would fewer and less flexible items of the other type. In making this choice, the investigator was fully aware of the discrepancy which so often appears between *opinion* and *overt action*. Yet, in the appraisal of *attitude*, opinion is perhaps as valid an indicator as is potential overt behavior.

The subject was to respond to each item by encircling a number representing the degree of his acceptance as follows: (1) strongly agree, (2) agree, (3) undecided or uncertain, (4) disagree, (5) strongly disagree. The following example illustrates a "strongly agree" reaction:

112. Immediate obedience is not always desirable. ① 2 3 4 5

An extensive survey of the literature was made to obtain a wealth of ideas concerning all phases of teacher-pupil relationship, both in and out of the classroom, to be used as source material for the building of items. These ideas, coupled with those from the writer's own experience, provided the basis for the compilation of a list of 378 opinion statements relating to the reaction of teachers to children and pupils and their behavior. Items expressing opinions concerning children in general were included along with those relating directly to pupils, as it was felt that a teacher's reaction to *children as such* is of significance in her attitude toward pupils. Items were constructed with children of elementary school age primarily in mind. It was believed that the Inventory eventually would prove most useful at this level.

3. *Classification of the Items*

On the assumption that the employment of some system of classification in item construction would contribute to a more adequate sampling of teacher attitude and behavior in relation to pupils, the opinion statements were classified under five categories which are presented here with brief explanatory comments and illustrative items:

(1) *Moral status.* This phrase is meant to cover those adult-child relations essentially characterized by adult recognition or non-recognition of child or pupil as an individual with his own inherent rights. Placed under this category were items that related to the status of children in the eyes of adults, especially as concerns their adherence to adult-imposed standards, moral and otherwise.

Example: Children "should be seen and not heard."

(2) *Discipline.* Under this caption were included those statements relating to problems of conduct in the classroom and elsewhere, and the methods employed in dealing with them.

Example: Pupils found writing notes should be severely punished.

(3) *Child knowledge.* Here were placed those statements relating to principles of child development and behavior ordinarily included in such subject fields as educational psychology, child psychology, and mental hygiene. More specifically, the content of these items related to the topics of ability, achievement, learning, motivation, and personality development.

Example: The boastful child is usually overconfident of his ability.

(4) *Educational principles.* Under this category were placed iems pertaining to educational principles and practices not *directly* relating to child development and behavior. Problems in education concerning method, philosophy, curriculum, and administration are touched upon here.

Example: Pupils should be required to do more studying at home.

(5) *Personal reactions of teacher.* Included under this category were statements directly relating to the teacher's personal disposition—her likes and dislikes, sources of irritation, etc.

Example: Without children life would be dull.

Consideration of the above system of classification will reveal a rough attempt to identify psychological factors in personality structure and function, as well as to make a differentiation relative to the ideational content of items. It was believed, for example, that the categories "Moral status" and "Personal reactions of teacher" related to responses that involved more of the affective nature and less of the cognitive, than did the other categories. The feeling component in these areas would seem to be stronger than the intellectual element. At least, the subject-matter that would make up the content of items under these two categories was less objective and less well established as ideational principles. However, the composition of psychological reactions to statements of opinion, as expressed in terms of affective and intellectual elements, is still essentially a matter of conjecture.

Added to this uncertainty is the problem of overlapping among the categories. As the classification of statements proceeded, the difficulty of the process became more and more apparent, with an increasing realization that a number of the items should be classified under two or more categories because of a very evident relevance to each of several topics. The following item, for example, was classified under the categories of "Moral status," "Discipline," and "Educational principles":

> Unquestioning obedience in a child is not desirable.

And, who is to say that this item, with possibly equal justification, could not have been placed under the two remaining categories of "Child knowledge" and "Personal reactions of teacher?"

Such considerations as the above made the whole problem of item classification very difficult to handle, especially on a logical and an a priori basis. The classification process was completed, however, as it was felt that it would lead to a more substantial body of Inventory content. Table 1 indicates the number of items in the final Inventory of 164 items that were classified under each of the five categories. A number of the items were placed in more than one division.

TABLE 1

NUMBER OF ITEMS IN FINAL INVENTORY OF 164 ITEMS CLASSIFIED UNDER EACH OF FIVE CATEGORIES

Category	No. of Items
Moral status	110
Discipline	43
Child knowledge	43
Educational principles	50
Personal reactions of teacher	43

4. *Two Forms of the Inventory*

To facilitate the selection of the best-stated and most discriminating items, two forms (A and B) of the *Teacher-Pupil Inventory* were constructed. Form B was made up of a second list of 378 items, or opinion statements, each of which (with the exception of 8 or 10 items) was stated in a different manner (usually in reverse form) from a correspondingly numbered related item in the first list (Form A). Samples of several items from Form A with correlative items from Form B are given below.

1. Most children are obedient. (Form A)
1. Most children tend to be disobedient. (Form B)

2. Pupils expect too much help from the teacher in getting their lessons. (Form A)
2. Teachers should give pupils more direct help with their lessons. (Form B)

3. The majority of children take their responsibilities seriously. (Form A)
3. Most children are irresponsible. (Form B)

In order to make Form A and Form B as equivalent as possible, approximately 50 per cent of the items in each form were so worded that agreement with these statements corresponded to the same end of the attitude continuum as did disagreement with the other 50 per cent of the items. These two kinds of statements were distributed at random throughout each form. Items representing the five categorical topics were also placed at random within each form. The number of items in each form (378) was made amply large to provide for the elimination of poor items.

B. ADMINISTRATION OF THE ORIGINAL INVENTORY

1. *Empirical Basis for Standardization*

Instead of using judges or experts to rate the items as to validity and quality, to establish a basis for standardization, and to devise a system of scoring, it was the plan of the investigation to arrive at these ends by *empirical* means. That is, the plan was to administer the Inventory to two groups of teachers (one group characterized as maintaining good relations with pupils, the other group as maintaining poor relations) to determine the *differences,* if any, between the groups in their reaction to the items. In other words, instead of logically determining in advance how certain teachers *ought* to react to the items and then seeing if they measure up to this standard, the method of empiricism involves the experimental discovery of how various teachers actually *do* react to the items, which finding then serves as a basis for the interpretation of all subsequent reactions.

The criterion used in the original selection of the two groups of teachers was the local school administrator's judgment concerning the teacher-pupil relationship status of his individual faculty members. In employing this criterion, the investigator fully realized that administrators often are not as aware as they should be of the actual nature of the relationship existing between the pupils and individual members of the teaching staff.

2. *Nature of the Teacher Population*

The teacher population to which the original *Teacher-Pupil Inventory* was administered comprised two groups of teachers, 100 in each group, differentiated on the basis of the local school administrator's judgment concerning the individual teacher's ability to maintain harmonious relations with pupils. This was the only point of differentiation that was controlled in the selection of the two groups. Each group was heterogeneous in nature as there was no attempt made to control such factors as age, sex, marital and parental status, nationality, grade level, subject taught, training, and experience. Little consideration was given these factors at this stage of the investigation, as the essential focus of attention here was the *difference* in reaction to items of the Inventory between the two groups of teachers who had been differentiated solely on the basis of relationship with pupils. The writer was aware of the possible influence of the other factors upon teachers' attitudes, but consideration of this problem was reserved until a later time.

TABLE 2

NUMBER OF SCHOOLS AND TEACHERS REPRESENTING DIFFERENT-SIZED COMMUNITIES AND DIFFERENT SCHOOL LEVELS FROM WHICH TWO GROUPS OF 100 TEACHERS EACH WERE DRAWN

Size of Community	Superior Group			Inferior Group		
	No. of Schools		No. of Teachers	No. of Schools		No. of Teachers
Cities with population of 100,000 and over	Sr. High	2	3	Sr. High	1	4
	Jr. High	1	1	Jr. High	1	2
	Elem.	9	14	Elem.	8	16
Cities with population between 1,000 and 100,000	Sr. High	9	21	Sr. High	8	20
	Jr. High	5	10	Jr. High	4	9
	Elem.	17	30	Elem.	18	33
County schools and towns with population of 1,000 and under	High	14*	9	High	11*	9
	Elem.		12	Elem.		7
		57	100		51	100

* Both elementary and secondary grades housed in one building.

An effort was made to contact teachers from schools of varying size and description so as to include individuals with varied teaching environments. However, the study was limited to teachers in the public schools at the elementary and high school levels.

The schools included were located in Pennsylvania and Ohio and ranged in size from one-room county schools to the elementary and high schools of one of our larger cities. The schools represented farming communities, industrial and mining regions, mercantile areas, and wealthy residential districts. None of the schools was extremely progressive or ultraconservative. In the investigator's judgment they were, in general, educational institutions average in philosophy and practice, probably tending toward the traditional and conservative type of school.

Table 2 indicates the extent to which different-sized communities were represented in making up the two groups of 100 teachers each. The representation, in each group, of the different school levels is also indicated. No effort was made to equate the two groups relative to the size of school system or to the grade level represented. Teachers at each extreme of the teacher-pupil relationship continuum were used irrespective of where they were found. Reference to Table 2, however, will reveal a close similarity in composition of the two groups.

3. *Personal and Descriptive Data*

By means of a Form for Personal and Descriptive Data which accompanied the Inventory, a record was kept, for each teacher, of the factors of age, sex, marital and parental status, nationality, grade level, subject taught, training, and experience. The teacher was also asked (*a*) to indicate whether or not he had had a course in mental hygiene, and (*b*) to rate himself relative to his liking for teaching. The name of the teacher was not requested. It was felt that anonymous responses would approximate more closely actual attitudes.

4. *Explanation and Directions for Completing the Inventory*

Explanation and instructions relative to completing the Inventory were as follows:

> *Explanation and Directions*
>
> There is considerable disagreement in educational circles about a variety of topics involving the relationships of teachers and pupils. These topics are still largely matters of opinion with little reënforcement from scientific evidence. It is the purpose of the Inventory to sample existing opinion in the area of teacher-pupil relationships. There are no right or wrong answers to the items of the Inventory. *Your own individual feeling* about the matter is what is wanted. A *frank* and *sincere* response to each item is obviously necessary if this study is to serve its purpose. Your responses will have no effect whatever upon your personal or professional standing. *Your name is not requested.*
>
> *Directions:* After reading the item carefully, indicate your response by encircling the number which most nearly represents your reaction to the topic in question. Think in terms of general situations rather than of individual cases. *Please respond to every item.* The scale is as follows:
>
> 1. Strongly agree
> 2. Agree
> 3. Undecided or uncertain
> 4. Disagree
> 5. Strongly disagree

5. *Distribution of the Inventories*

Before the teachers were approached, permission to carry on the investigation in the particular school system and building was obtained from the superintendent and principal, respectively, and their cooperation was assured. The purpose and methodology of the study were carefully explained to them. The school official feeling most qualified (in most cases this was the principal) was asked to list, if possible, several teachers on his staff whom he considered quite superior in their relationships with pupils and several at the opposite extreme, whom he considered quite inferior in this respect. In making this selection, he was to consider the following criteria with reference to each teacher:

(1) Ability to win the affection of his pupils.
(2) Fondness for, and understanding of, children.
(3) Ability to maintain a desirable form of discipline.

Although the characteristics above are not necessarily correlative, the school official generally experienced little difficulty in making the selection desired. It was made clear to him that, among a number of factors that go to make up the professional requirements of a teacher, relationship with pupils was the only one that he was to consider in his selection. Such factors as intelligence, teaching methods, and knowledge of subject-matter were not to be used as differentiating criteria.

The investigator presented a copy of Form A of the original Inventory to each of the 200 teachers in person.[1] Each was asked if he would cooperate in a research project by completing the Inventory sometime within the next several weeks, at the end of which time the writer would return for it. Each teacher was assured that no one but the writer would have access to the completed Inventory. It was made clear that the project was the writer's only, and that the school system or administration had had no part in the instigation of the study. Although there were a few teachers who did not wish to participate, cooperation, for the most part, was excellent. For purposes of identification, a record was kept of the number of the copy that was given to each teacher.[2]

---

[1] Although consuming much time and effort, this individual, personal contact was considered essential to the success of the investigation.

[2] This number had been placed on the Inventory quite inconspicuously (immediately following the form designation; e.g., Form A6).

Thus, from none to five or six teachers at each extreme (depending, among other factors, upon the size of the teaching staff) were located in each of sixty-seven buildings that were visited in fourteen school systems. In this manner there were procured 100 subjects in each of two extreme groups—one group composed of teachers who were considerably above average in their ability to get along with pupils (to be known as the "superior" group), and the other group consisting of teachers considerably below average in this ability (to be known as the "inferior" group).

After a month or so had elapsed, the writer revisited the sixty-seven schools and again saw, individually, each of the 200 teachers. At this time Form A of the Inventory was taken up and a copy of Form B, with a return envelope, was left with each teacher. For purposes of identification care was taken that the number on the copy of Form B given to any one individual corresponded with that on his copy of Form A.

C. COMPARISON OF RESPONSES AND SELECTION OF ITEMS

As soon as the Inventories were completed and returned, the response of each of the 200 teachers to each of the 378 items in each of the two forms of the Inventory was carefully recorded. The purpose of this work was to *compare* the responses of the two groups of teachers to each of the items in order to determine possible *differences* in reaction. Items that did not differentiate the groups could then be discarded and a scoring system devised for the remaining items on the basis of the obtained differential responses.

1. *Comparison of Responses to the Items*

The frequency of each of the five responses (1, 2, 3, 4, 5) for each item was obtained, separately, for each group. As there were 100 individuals in each group, the frequencies also took the form of percentages. As illustrated in Table 3, a table showing the pattern of response was constructed for each of the 756 items (Forms A and B).

The response pattern in Table 3 is interpreted as follows: 174, or 87 per cent, of the 200 teachers are in agreement with this item (including both the "strongly agree" and "agree" responses). Thirty-four per cent of the superior group "strongly agree"; 18 per cent of the inferior group "strongly agree"; 3 per cent of the superior group "disagree," whereas 13 per cent of the inferior group "disagree"; the same percentage of each group are "undecided" (4 per cent); the same percentage "strongly disagree" (1 per cent). It is of interest to note that a larger percentage of the inferior group (64 per cent) than of the superior group (58 per cent) "agree" to the item. This apparent discrepancy is probably explained by the fact that a relatively larger proportion of the superior group is found in the "strongly agree" column. This ex-

TABLE 3

DISTRIBUTION OF RESPONSES OF THE SUPERIOR AND INFERIOR GROUPS OF TEACHERS TO ITEM 1 (FORM A)

1. Most children are obedient.

Group	Strongly Agree *1*	Agree *2*	Undecided *3*	Disagree *4*	Strongly Disagree *5*
Superior	34	58	4	3	1
Inferior	18	64	4	13	1
Difference	16	−6	0	−10	0

TABLE 4

Distribution of Responses of the Superior and Inferior Groups of Teachers to Item 243 (Form A)

243. If the teacher laughs with the pupils in amusing classroom situations, the class tends to get out of control.

Group	Strongly Agree 1	Agree 2	Undecided 3	Disagree 4	Strongly Disagree 5
Superior	1	0	1	55	43
Inferior	1	4	6	67	22
Difference	0	−4	−5	−12	21

planation became more and more plausible as the reactions to other items were studied. Consider the response pattern in Table 4 as a second example. Over 90 per cent of the teachers do not agree with this item. The most differentiating response between the groups is No. 5 ("strongly disagree"). It seems reasonable to interpret the relatively high percentage (67 per cent) of the inferior group found under Response 4 ("disagree") as a general disagreement on the part of the teachers of this group concerning the item, but as an apparent unwillingness to go to the extreme in their response. In other words, one might characterize their reaction as one of *nominal* disagreement, or as representing the most widely accepted opinion, whereas the reaction of the superior teachers is strong enough to indicate an attitude of sufficient potency to effectively influence their daily professional work. With a number of items, in this manner, the strength of differentiation lay in the difference in response of the groups relative to the two degrees of reaction at either end of the attitude continuum.

The responses indicated in Table 5 assume a different pattern. The responses for this item

TABLE 5

Distribution of Responses of the Superior and Inferior Groups of Teachers to Item 290 (Form A)

290. Most pupils don't appreciate what a teacher does for them.

Group	Strongly Agree 1	Agree 2	Undecided 3	Disagree 4	Strongly Disagree 5
Superior	0	38	16	39	7
Inferior	7	49	15	29	0
Difference	−7	−11	1	10	7

TABLE 6

Distribution of Responses of the Superior and Inferior Groups of Teachers to Item 164 (Form B)

164. As a rule pupils do not have much use for teachers.

Group	Strongly Agree 1	Agree 2	Undecided 3	Disagree 4	Strongly Disagree 5
Superior	1	4	3	67	25
Inferior	1	7	15	61	16
Difference	0	−3	−12	6	9

assume a rather striking pattern of symmetry. A larger proportion of the inferior group, however, are in agreement with the item. It is of interest to note that the same number (7) of inferior teachers "strongly agreed" to the item as superior teachers who "strongly disagreed." There is also a remarkably close agreement between the groups relative to Response No. 3 ("undecided"). This similarity of response in the "undecided" category is characteristic of a number of the items.

There are other items, however, with which considerable differentiation between the groups occurs in the "undecided" response. Table 6 illustrates an item in which most of the differentiation occurs in this category. Although with this item 84.5 per cent of the 200 teachers are in disagreement, *undecidedness* seems to characterize the inferior group more than it does the other group.

2. *Selection of Items for the Final Inventory*

Only 164 of the original 756 items were chosen for use in the final Inventory[1], the remaining items being rejected. The criteria employed in the selection and rejection of items were as follows:

(1) Was the item adequate in differentiating the two groups of teachers?

(2) Was the item ambiguous in meaning, lacking in clearness, or poorly stated?

(3) Did the content of an item duplicate that of another item that had been selected?

(4) Did the item show a response pattern that was difficult to interpret?

The statistic, chi-square, was employed as an objective measure of the degree of differentiation between two response patterns. The following formula (7, pp. 385-386) was used to calculate the contribution to chi-square of any given pair of responses:

$$\frac{1}{a + a'}(aN' - a'N)^2 \text{ where,}$$

a and a' represent the obtained column entries and N and N' represent the corresponding row totals.

As a valid and practical measure of the degree of differentiation between two response patterns, chi-square was found to have its limitations. Large chi-square values were sometimes the result of differences in response that were found in two categories only (of which one was often the "undecided" response) and of differences in response that were irregularly distributed throughout the attitude continuum, thereby making interpretation difficult. A few items with relatively high chi-square values were, consequently, rejected for use in the final Inventory. On the other hand, there were a few items with relatively low chi-square values which were selected for use because of their symmetrical response patterns or because their inclusion in the Inventory would lead to a richer and more diversified body of content for the instrument.

The value of chi-square proved to be an effective criterion for item selection *provided the remaining three criteria were satisfied.* Generally speaking, an item (even though possessing a high chi-square value) was discarded if it did not meet the demands of all three remaining criteria.

Of the 164 items chosen for use in the final Inventory:

79 (48 per cent) met the probability level of .05 ($\chi^2 = 9.488$)
112 (68 per cent) met the probability level of .10 ($\chi^2 = 7.779$)
134 (82 per cent) met the probability level of .20 ($\chi^2 = 5.989$)
146 (89 per cent) met the probability level of .30 ($\chi^2 = 4.878$)
157 (96 per cent) met the probability level of .50 ($\chi^2 = 3.357$)

It was in the original plan of the investigation to so construct the *Teacher-Pupil Inventory* that, in its final state, it would contain from 150 to 175 items in each of two forms. As analysis of items progressed, however, it became increasingly apparent that the final instrument would be more satisfactory if it consisted of one form with a more select group of items than of two forms with inferior items. The revised Inventory, then, issued in one form containing 164 of the most select items.

D. THE ORIGINAL SCORING SYSTEM FOR THE INVENTORY

The simplified scoring system of Likert (use of consecutive numerical values 1, 2, 3, 4, 5, assigned to the several responses) had been contemplated for use with the present instrument. When responses to the items were analyzed, however, the idea of using this technique was abandoned. It became increasingly evident that, in the great majority of cases, the responses for each item were not normally distributed. Rough inspection revealed tendencies toward distributions bi-modal in character, or with a piling up of responses at one or the other end of the attitude continuum. Furthermore, no assumption was to be made relative to uniformly—or non-uniformly—spaced "steps" along the attitude continuum.

1. *Procedure Used in Calculation of Weights for Item Responses*

Consideration of such factors as the above, coupled with the desire to make the investigation as empirical as possible, led to the adoption of a procedure for weighting the several responses for each item in accordance with the *differentiation* empirically established between the two groups of teachers studied. This was the method employed by Strong (17) in weighting the responses to the items of his Vocational Interest Inventory. Strong's formula, contributed by T. L. Kelley (17, pp. 603-615), was applied to the construction of the scoring key for the *Teacher-Pupil Inventory*. The range of weights used extended from plus 4 to minus 4. Weights beyond these limits were valued at plus 4 and minus 4, respectively. For each of the five responses associated with each of the 164 items in the revised Inventory the formula was used to compute the weight to be assigned to that response. The value of W obtained by use of the formula is multiplied by 100, and the nearest whole number is used.

E. THE SIMPLIFIED SCORING SYSTEM FOR THE INVENTORY

The system of scoring indicated above, with weights ranging from plus 4 to minus 4 assigned to the several responses, necessitates the rather cumbersome process of adding algebraically a fairly long list of positive and negative numbers. Later in the study it was found that results produced by a scoring scheme which allotted plus 1 to each circled response with a positive weight (disregarding all circled responses with negative and zero weights) correlated .973 with results attained by the longer method. This simpler scoring method was therefore later adopted.

F. COMPARISON OF INVENTORY SCORES OF THE SUPERIOR AND INFERIOR GROUPS OF TEACHERS

After the original scoring key had been established, it was used in the scoring of the 200 inventories received from the two groups of teachers. Forms A and B were now combined, and, in the scoring, consideration was limited to those 164 items which had been selected for the final form of the Inventory. Results obtained in comparing the scores of the superior and inferior groups of teachers

TABLE 7

RESULTS OBTAINED IN THE COMPARISON OF INVENTORY SCORES OF THE SUPERIOR AND THE INFERIOR GROUPS OF TEACHERS

Statistic	Superior Group	Inferior Group
N	100	100
Range	324—(−99)	213—(−297)
M	131.0	−32.0
S.D.	102.91	112.46
σ_M	10.19	11.18
Mdn	130.0	−41.5
σD_M	15.13	
D_M	163.01	
C.R.	10.8	

are presented in Table 7. It may be seen that, despite some overlapping, there is a highly significant difference between the means of the two groups of teachers (critical ratio: 10.8). Although not shown in the table, 10 of the scores of the inferior group were above the mean of the superior group, and only 4 of the scores of the latter group were below the mean of the inferior group. Also, it is of interest to note that, in the superior group, the highest score (324) was obtained by a Dean of Women in a high school and a score fifth from the top (295) was made by a junior high school teacher who also was serving in the capacity of visiting teacher. These were the only two individuals in the sample occupying such unique professional status.

1. *Distribution of Inventory Scores for the Superior and Inferior Groups of Teachers*

The statistics g_1 and g_2, suggested by Fisher (15) as appropriate tests for the normality of a distribution, were calculated for each of the two sets of scores. The quantity g_1 is a measure of skewness; the closer its value to zero, the more symmetrical the distribution. The negative value of g_1 obtained for the superior group, as shown in Table 8, indicates a slight asymmetry with an excess of scores larger than the mean. The positive value of g_1 found for the inferior group indicates an excess of scores smaller than the mean. However, each of the g_1 values is too small to be significant.

TABLE 8

VALUE AND SIGNIFICANCE OF g_1 AND g_2 FOR THE DISTRIBUTIONS OF INVENTORY SCORES MADE BY THE SUPERIOR AND INFERIOR GROUPS OF TEACHERS

Group	g_1	t_{g_1}	Probability of t_{g_1}	g_2	t_{g_2}	Probability of t_{g_2}
Superior	−.0910	.3789	$>.05$	−.8342	1.744	$>.05$
Inferior	.0927	.3860	$>.05$	−.4472	.9349	$>.05$

The statistic g_2 is a measure of kurtosis. If its value is zero, normality is indicated so far as this measure is concerned. The negative values of g_2 obtained for the two distributions indicate flat-topped distribution curves. However, each of the g_2 values is too small to be significant.

III. THE VALIDITY AND RELIABILITY OF THE TEACHER-PUPIL INVENTORY

Although subjective in nature, the process of composing the items contributed to the validity of the instrument. In a more objective manner the administration of the Inventory to two groups of teachers at extreme ends of an attitude-toward-pupils continuum constituted a validating process *in establishing the validity of individual items.* Other criteria being satisfied, only items that served to *differentiate* the groups were selected for use in the final or revised instrument. The statistic, chi-square, served as the measure of the degree of differentiation.

It now remained to establish the validity of the Inventory as a whole.

A. Determining the Validity of the Inventory

1. *Plan of Procedure*

The plan of determining the validity of the Inventory essentially involved (*a*) its administration to a new sample of 100 teachers, (*b*) the obtaining of ratings on these teachers with reference to their relationships with pupils, and (*c*) the correlation of the teachers' scores on the Inventory with the ratings. These ratings were to be obtained from each of three sources: (*a*) principals under whom the teachers were serving, (*b*) classroom observation on the part of the investigator, and (*c*) attitudes of pupils toward each individual teacher.

As a further check on the validity of the items of the Inventory, out of the new distribution of 100 teachers the item response patterns of the twenty with the highest Inventory scores and the response patterns of the twenty with the lowest scores were to be compared, respectively, with the patterns of item response characterizing the original groups of superior and inferior teachers. Also, the responses of the twenty teachers with the highest scores were to be compared with the responses of the twenty with the lowest scores. The statistic, chi-square, was to be employed as the objective means of comparison.

2. *Administration of the Revised Inventory*

a. The nature of the teacher population. Inasmuch as the original sample of 200 teachers represented the two extremes in teacher-pupil relationship, it was deemed desirable to include in the new sample 100 teachers *unselected* with reference to their relationship with pupils. In designation, this group will be known as the "unselected" group.

The 100 teachers who were to serve as subjects were fourth-, fifth-, and sixth-grade teachers in three different school systems, two of which had been used in the earlier phase of the investigation. In these two systems, however, buildings different from those used previously were visited at this time. The one condition under which the 100 teachers were selected, aside from their status as public school employees, was that they have teaching contact with children in the fourth, fifth, or sixth grade at the time of testing. It was immaterial as to whether they were teaching one or more of these grades. Table 9 indicates the number of cities, schools, and teachers representing the different-sized communities from which the unselected group of 100 teachers was drawn.

TABLE 9

Number of Cities, Schools, and Teachers Representing Different-Sized Communities from Which the Unselected Group of 100 Teachers Was Drawn

Community Size	No. of Cities	No. of Schools	No. of Teachers
Cities with population of 100,000 and over	1	5	46
Cities with population between 1,000 and 100,000	2	13	54
Total	3	18	100

It was planned to use teachers of the elementary intermediate grades for two fundamental reasons: (*a*) Since the inception of the study, it had been the investigator's conviction that the prospective Inventory would prove of particular value at this important rung of the educational ladder. For where else than in the elementary grades is the matter of teacher-pupil relationship of such peculiar importance? Consequently, items were constructed with children of elementary school age primarily in mind. (*b*) It was believed that the attitudes of pupils toward their teachers (an important validating device in the study) could be obtained more effectively at the intermediate grade level. Pupils of grades four, five, and six would be of sufficient maturity to respond adequately to simple items in a questionnaire. Furthermore, it was felt that responses obtained from these pupils would be of a more ingenuous and less sophisticated nature than those that would be procured from pupils more advanced in educational achievement, who have far too often, become comfortably adapted to the conventional school situation.

b. Mechanics of administration. As in the previous visitation of schools, the initial step at this time was to explain the nature of the investigation to the administrative head of the particular school system and obtain the assurance of his cooperation in the study. After then interviewing the principals of the several buildings that were to be used and explaining the purpose and method of the investigation, the writer contacted all teachers at the fourth-, fifth-, and sixth-grade levels personally and individually. Unless the teacher preferred not to participate in the study, a copy of the revised Inventory, together with a stamped return envelope, was left with him.[1]

3. *Obtaining the Ratings of the Teachers*

Care was taken that no attempt was made to secure further data until all copies of the Inventory distributed to the teachers had been completed and returned. This precaution was considered essential, for any awareness on the part of the teacher that he was to be rated probably would have influenced his responses to the items in the Inventory. On the other hand, to avoid bias in the mind of the investigator in his ratings of teacher behavior, scoring of the Inventories was deferred until all data relative to rating of the teachers had been gathered.

a. Ratings by principal. A rating scale was devised on which a principal was to rate each of the teacher-subjects on his staff with reference to relationship with pupils. Ratings were to be obtained for each teacher on the following aspects of this relationship:

(1) Disciplinary ability.
(2) "Personnel" vs. "subject-matter" point of view.
(3) Attitude toward children.
(4) Understanding of pupil behavior problems.
(5) Personality adjustment.
(6) Attitude of pupils toward this teacher.

It would seem that the above items are important factors in the personal interaction between teacher and pupil. This personal interaction is the aspect of the whole teacher-pupil relationship pattern that is of particular interest in the present study. The topics mentioned represent several basic areas of teacher-pupil relationship from which a number of the items of the Inventory were drawn.

---

[1] The Inventory copies again had been numbered to make identification possible.

The discipline that a teacher maintains in the classroom essentially is determined by the personal relationship that exists between him and the pupils. Whether it is the tyrannical teacher who "rules with an iron hand" and instills fear into his pupils or the supine teacher who is unable to "keep order" and whose pupils have no respect for him, the problem is principally one of affective and personal relationships.

The personal interaction between teacher and pupil probably is influenced by the teacher's own philosophy of education. At one extreme he may look upon the child merely as a passive receptacle for knowledge and dogmatically cling to "blanket" standards of achievement. Or, on the other hand, he may regard the mastery of subject-matter as secondary and actively strive to meet the varied needs of the child's developing personality.

Item number 3 in the Rating Scale, concerned with the teacher's attitude toward children, goes right to the heart of the problem with which the Inventory was designed to cope. Does the teacher look upon the child as a sinister and inferior being whose greatest virtue is to obey his elders? Or, at the other extreme, does the teacher naturally love children, and, sympathetically and understandingly, appreciate their individual problems of growing up?

Closely related to a teacher's attitude toward children is his understanding of child behavior problems. In his reactions toward the behavior of children does he indicate an understanding of, and an attempt to apply, the principles voiced by mental hygienists? For example, does he consider the behavior of the shy, withdrawing child as probably more serious than the aggressive transgression against authority? Does he actively seek to determine the causes behind misconduct, or is he satisfied with the consideration of symptoms only?

Of unquestioned influence in a teacher's personal interaction with pupils is the status of his own personality adjustment. It is a strong conviction of the writer that teachers with numerous disciplinary problems, teachers who are irritated easily by children's actions, and teachers who are often uneasy in the presence of children are beset with personality difficulties of their own which undermine their effectiveness in the schoolroom.

The attitude of the pupils toward a teacher is generally sensed by the principal. Especially is this true with a very popular teacher or with one who is regarded as an enemy. As indicated elsewhere in this paper, the pupils' liking or disliking for a teacher is considered of basic significance in teacher-pupil relationship.

When all of the inventories had been returned, copies of the rating scale described above, together with mailing facilities for their return, were left with each of the twelve principals in the eighteen buildings visited in this part of the investigation. The discrepancy between the numbers of principals and buildings was due to the fact that, in two of the school systems, each principal had charge of two or three buildings.

The mean of the principals' ratings was 5.62 with a range of scores extending from 14 (perfect score) to −9 (−14 being the lowest possible score).

b. Classroom observation. The next process in determining the validity of the Inventory involved the writer's observation of the classroom behavior of each of the 100 teachers relative to his relationship with pupils. As an aid in this observation, use was made of Baxter's *Rating Scale of the Teacher's Personal Effectiveness* (1), modified so as to meet more adequately the needs of the present study. Baxter's Scale was devised with the principal object of evaluating a teacher's personal effectiveness in the classroom by considering the pupils' responses to his conduct. This scale was modified for use in the present investigation by (*a*) omitting the responses of the pupils, (*b*) omitting two of Baxter's items which were somewhat irrelevant to the present study, (*c*) adding an item relating to the teacher's disciplinary ability, and (*d*) slightly altering the wording of Baxter's items in several places. It was believed that such modification would facilitate the evaluation of each teacher's personal reactions toward the pupils.

The writer spent from twenty minutes to one hour in the classroom of each teacher to observe his effectiveness in the following:

(1) Maintaining discipline.
(2) Creating a friendly classroom atmosphere.
(3) Establishing a feeling of security.
(4) Exerting a stabilizing influence.
(5) Developing pupil self-reliance.

The significance of the first two points above in teacher-pupil relationships is evident. Also important in the personal interaction between teacher and pupil is the feeling of security that a teacher is able to establish in his pupils. He may be encouraging and able to inspire confidence, or he may be intolerant of mistakes and definitely lacking in appreciation of the pupils' efforts.

Effectiveness in "exerting a stabilizing influence" essentially depends upon the teacher's own personality adjustment. The importance of this problem in teacher-pupil relationship cannot be over-emphasized.

A teacher's skill in developing pupil self-reliance is an important factor in his personal relationships with pupils. On the one hand, he may encourage pupil decision; on the other, he may demand rigid conformity.

The rating of each teacher relative to the above points provided a numerical estimate of his personal effectiveness with the pupils which served as one criterion in establishing the validity of the Inventory. The mean of the classroom observation scores was 3.60, and the range of scores extended from 10 to −9 (possible range from 10 to −10).

c. Attitudes of pupils. The third phase in determining the validity of the Inventory involved procuring the reactions of pupils to each of the 100 teachers. Of the three validating criteria, that considered of most importance was the attitudes of pupils toward their teachers. Who is in a better position to decide what a pupil thinks about his teacher than the pupil himself who is with him so much of each school day? Or, isn't what a pupil thinks about the teacher of much importance? A positive answer to this question represents a basic presupposition of the present study. Even though children are immature, when gauged by the standards of adults, and even though their opinions of teachers may change considerably as they grow older, it would seem that it is their *present* attitude, infantile as it may seem, that should be of greatest significance to the educator.

The measuring device constructed to procure an estimation of the pupils' attitudes toward the teacher took the form of a 50-item questionnaire entitled *My Teacher.*[2] The items were simple questions adapted to the mental maturity and experience of fourth-, fifth-, and sixth-grade children, that could be answered simply by underlining "Yes," "No," or "?."

The questions were formulated by the writer on the basis of ideas gleaned from the literature on teacher-pupil relationships. The content of the questions related to the teacher's disposition, his treatment of pupils, his teaching, his interest in children's activities, his status in the pupils' esteem, his sense of humor, and a number of other topics within the realm of teacher-pupil relationship. The list of questions was made long enough to exhaust the source material for item content, yet short enough that completion of the questionnaire would not tax the interest of a pupil in the intermediate grades. The following questions are cited as examples:

4. Does this teacher scold the pupils a lot? Yes No ?
6. Does this teacher explain the school work so that you can understand it? Yes No ?
11. Is this teacher often "bossy"? Yes No ?

---

[2] A copy of *My Teacher* may be found in Appendix D in the writer's Ph.D. dissertation (10) on file in the University of Minnesota Library.

38. Does this teacher give you a chance to ask questions? Yes No ?
49. Do you like this teacher? Yes No ?

Space was provided on the questionnaire for the pupil to comment on (*a*) Why I Like This Teacher, and (*b*) Why I Don't Like This Teacher. The pupil was also asked to indicate the grade or mark which he usually received from this teacher.

As a safeguard against disturbing morale, classroom observation on the part of the writer was completed before pupil reactions were obtained.

In procuring these reactions, the general plan was to administer the questionnaire, in person, to groups of about 25 fourth-, fifth-, or sixth-grade pupils. As these grades were departmentalized, in whole or in part, in two of the three school systems used in this phase of the study, the majority of the pupils had classroom contact with several teachers during the school day. For practical reasons, in some cases a group of pupils rated only one teacher, whereas in other cases, a group rated two, or possibly three, teachers. However, only one teacher was rated during any one administration of the questionnaire. A condition that had to be met was that any teacher being rated by a group must have teaching contact, that school year, with the pupils in the group.

All the data for the investigation were gathered during the second half of the school year, at which time it was felt that teachers and pupils had had ample opportunity for mutual acquaintance.

To make conditions as uniform as possible, pupils were not asked to rate their home-room teachers. There is reason to believe that a home-room teacher occupies a somewhat different status in the thoughts and feeling of her home-room pupils from that of a teacher not serving in this capacity.

Each administration of the questionnaire (there were 100 separate administrations in all) required from fifteen to twenty minutes. Groups ranged in size from twenty to forty pupils, usually numbering about twenty-five. In some buildings it was the wish of the administrator that the writer go from room to room and use the whole class for each separate rating. In others it was considered desirable that the rating be done in one room. In the latter situation twenty-five pupils were selected at random from one of the intermediate grades by the principal and directed to go to the room where the rating was being done.

The teacher who was being rated was never present during the administration of the questionnaire. In fact, no one was present other than the pupils doing the rating and the writer.

The principal made the following points clear to each group of pupils: (*a*) No one, including the teachers and the principal, would ever know how they (the individual pupils) answered the questions, for no one, except the writer, would ever see the papers; (*b*) the way the questions were answered would have no effect whatever upon their grades or school standing. In the interest of morale, the writer had also requested each principal to ask the pupils not to discuss the questions afterwards with teachers or other pupils. After these remarks, the principal left the room. The writer then stressed the desirability of honest answers, and re-ëmphasized the fact that he would be the only one that would ever see the papers, and that even he would not know how any individual pupil answered the questions as *no names were to be signed*. The writer then read the questions over with the pupils, making explanatory comments, and answering any questions that were asked.

The mean of the distribution of scores on the pupils' questionnaires was 20.0, the score range extending from 41 to −17 (possible range from 49 to −49).

d. The reliability of the ratings. The reliability of ratings is usually determined by correlating the ratings made by two comparable judges. In the present study, the use of this procedure did not appear feasible or practicable, particularly in connection with the ratings made by the principals and the writer. The method presupposes use of the same scale by both judges, whereas the rating scale used by the principals was different from the one used by the writer. Different scales were purposely employed. It was believed that a more extended and diversified list of traits to be rated, made possible by two *different* scales, would contribute definitely toward the validity of the ratings. Even if the same scale had been used by both principal and writer, correlation of the two ratings in determining the reliability of the instrument would have been questionable procedure in that the two judges would not have been "comparable" because of differences in background, mental set, and knowledge of the teacher being rated. The same problem would have arisen in any attempt to employ a third party to do the rating in the capacity of a second judge.

The reliability of the ratings made by the principals and the writer could have been determined by correlating two ratings made by the same individual at different times. It has been contended, however, that two ratings by the same individual are no better than one (18, p. 93). This fact coupled with the writer's hesitancy to impose upon the good will of school officials participating in the study led to the rejection of such a procedure.

(1) *Use of split-half technique.* The method used to determine the reliability of the rating scales was the split-half technique of correlating the odd items against the even items. Use of this method with rating scales is not very satisfactory because of the small number of items, the heterogeneity of their content, and the probable operation of the halo effect. The halo effect would tend to raise the coefficient of reliability, while at the same time lowering the validity of the scale. Regardless of its disadvantages, the split-half method was employed, and use was made of the Spearman-Brown formula to estimate the reliability of the whole scale from the calculated reliability of the half.

(2) *Reliability of the scale used by the principal.* By use of the split-half technique and the Spearman-Brown formula, the reliability coefficient of .87 was obtained for the principal's scale. Partly responsible for this result, undoubtedly, were the halo effect and the so-called "logical error"—the tendency for a judge to give similar ratings in traits that seem logically related in his own thinking.

(3) *Reliability of the scale used by the writer.* A reliability coefficient of .92 for the scale used by the writer in classroom observation was obtained by means of the split-half method and the Spearman-Brown formula. The high value probably is due in part to such influences as the halo effect and the logical error.

(4) *Reliability of the questionnaire used by the pupils.* The coefficient of reliability was determined for the pupils' questionnaire, *My Teacher,* by correlating the odd items with the even items from 200 papers. These 200 papers were chosen by selecting, at random, two papers from each of the 100 piles of questionnaires relating to each of the 100 teachers. The result was found to be .94 after the Spearman-Brown formula was employed. It would seem that the halo effect and logical error would be especially potent in their influence upon ratings made by elementary school children. Too, the failure of the split-half method to take account of the fluctuational changes in individuals from day to day results in an estimate of reliability that is probably too high from the standpoint of practical testing.

Inasmuch as from twenty to forty pupils rated each teacher it was possible to compute the reliability of the instrument when used by more than one rater. Twenty papers chosen at random from each of the 100 piles (one pile for each of the 100 teachers) were divided into two piles of ten papers each. After finding the mean score for each pile of ten papers, the results were used in correlating the mean of ten ratings with the mean of the other ten ratings for each of the 100 teachers. The correlation coefficient was found to be .83. This value represents the reliability of the instrument for ten raters. Substituting 2.5 in the Spearman-Brown formula, the reliability coefficient, .93, was obtained as the estimate of reliability of the instrument for twenty-five raters. The value for twenty-five raters was determined inasmuch as the majority of teachers were rated by groups of twenty-five pupils each.

It would seem that the value, .93, obtained for twenty-five raters, is of greater significance in the practical use of the questionnaire than the value, .94, which was the estimated reliability of the instrument as such when used by one rater. The instrument would ordinarily be used by a class of pupils rather than by a single pupil. Furthermore, the value of the reliability coefficient determined on either ten or twenty-five raters would appear to be closer to the true reliability sought when one considers the errors in rating previously discussed in connection with the split-half technique.

4. *Comparisons Among Ratings and Inventory Scores*

a. The distribution of Inventory scores for the unselected group of teachers. It is of interest to consider the distribution of Inventory scores for the unselected group of teachers in relation to the distribution of scores for the two original groups of teachers. Table 10 indicates the range of scores, mean, standard deviation, and median for each of the three distributions. The data are based upon the Inventory scores obtained by the original scoring process.

TABLE 10

RESULTS OBTAINED IN THE COMPARISON OF INVENTORY SCORES FOR THE SUPERIOR, UNSELECTED, AND INFERIOR GROUPS OF TEACHERS

Statistic	Superior Group	Unselected Group	Inferior Group
N	100	100	100
Range	324—(−99)	254—(−176)	213—(−297)
M	131.0	77.6	−32.0
S.D.	102.91	103.49	112.46
σ_M	10.19	10.35	11.18
Mdn	130.0	85.5	−41.5

As one would expect, both measures of central tendency (mean and median) for the unselected group lie between the central tendency values obtained for the superior and inferior groups. However, the obtained mean (77.6) for the unselected group is somewhat larger than the average (49.5) of the means for the two extreme groups. The same holds true for the obtained median (85.5) as compared with the average (44.2) of the other two medians. These differences may be due to one or more of such factors as (*a*) the limitation of the unselected group to teachers of the fourth, fifth, and sixth grades (seventh- and eighth-grade teachers in original samples were lowest); (*b*) the possibility that, of the two original groups of teachers, the inferior group was farther below the

TABLE 11

VALUE AND SIGNIFICANCE OF g_1 AND g_2 FOR THE DISTRIBUTION OF INVENTORY SCORES MADE BY THE UNSELECTED GROUP OF TEACHERS

Group	g_1	t_{g_2}	Probability of t_{g_1}	g_2	t_{g_2}	Probability of t_{g_2}
Unselected	−.37845	1.576	>.05	−.51469	1.076	>.05

parametric mean of an unselected population than the superior group was above it (weak teachers possibly more easily identified by principals than strong teachers); (*c*) the possibility that the unselected group was actually composed of better-than-average teachers; (*d*) chance errors.

It should also be observed that the range of the scores for the unselected group is from 254 to −176, whereas the total range for the extreme groups was from 324 to −297. The mean of the scores obtained by the simplified scoring system is 88.4, and the range extends from 120 to 38.

The distribution of Inventory scores for the unselected group was tested for normality as were the two original distributions. The statistics g_1 and g_2 were again employed. Table 11 presents the results obtained. The negative value of g_1 indicates an excess of scores larger than the mean or a mean smaller than the median. The value of g_1 does not differ significantly from zero, however, and asymmetry is not indicated. Although the negative value of g_2 indicates a tendency toward a flat-topped distribution curve, the value of g_2 is too small to be significant. As with each of the two original distributions, there is little evidence of departure from normality in the distribution of scores for the unselected group of teachers.

b. Correlations between Inventory scores and rating scale scores. Validity coefficients were found by computing correlations between scores on the revised Inventory and scores on each of the teacher rating devices. Table 12 indicates the coefficients obtained in comparing the Inventory results (both original and simplified scoring) with results of the ratings considered separately and in combination. Results obtained in correlating the ratings with each other are also presented. The single validating criterion showing the highest correlation (.486) with the Inventory is the writer's observation of the teacher's classroom behavior, whereas the criterion showing the lowest correlation (.434) is the principals' ratings. Between these values is found the correlation (.452) of the Inventory with the attitudes of the pupils. The fact that, of the three criteria, the

TABLE 12

CORRELATIONS OBTAINED BETWEEN INVENTORY SCORES (ORIGINAL AND SIMPLIFIED SCORING) AND RATINGS CONSIDERED SEPARATELY AND IN COMBINATION, TOGETHER WITH CORRELATIONS OBTAINED BETWEEN THE RATINGS, FOR THE UNSELECTED GROUP OF 100 TEACHERS

	Original Scoring		Simplified Sdoring	
	r*	σ_r	r*	σ_r
Inventory and Principals' Ratings	.434	.082	.445	.081
Inventory and Classroom Observations	.486	.077	.485	.077
Inventory and Pupils' Attitudes	.452	.080	.456	.080
Inventory and Three Validating Criteria Combined	.594 .595**	.065 .065	.596	.065
Inventory and Combination of Classroom Observations and Principals' Ratings	.536	.072	.544	.071
Inventory and Combination of Classroom Observations and Pupils' Attitudes	.576	.067	.579	.067
Inventory and Combination of Principals' Ratings and Pupils' Attitudes	.534	.072	.543	.071
Classroom Observations and Principals' Ratings	.480	.077		
Classroom Observations and Pupils' Attitudes	.326	.090		
Principals' Ratings and Pupils' Attitudes	.389	.085		

* These obtained r's are all statistically significant at the 1 per cent level (8, p. 324).
** Multiple correlation coefficient.

principals' ratings show the lowest correlation is rather interesting, inasmuch as the judgments of principals formed the criterion on which the selection of the original standardization groups was made.

It was believed that the three validating criteria combined, or two of them in the several possible combinations, would prove of greater value in establishing the validity of the Inventory than any one taken alone. It seemed reasonable to suppose that the combined judgment of several persons or groups relative to a teacher's behavior is of greater worth than the judgment of any one single agency. To make a combination of measures possible, all raw scores were converted into standard *T*-scores.

The correlation between the Inventory scores and the three validating criteria combined was found to be .594 (.596 with simplified scoring). When several criteria are thus considered together, extremes in one or two of them are counterbalanced by more moderate ratings in those remaining. A "tempering" or "leveling" process takes place which, apparently, tends to bring the resulting score into closer relationship with the score on the Inventory.

A multiple correlation of .595 between the Inventory scores and the three criteria was also obtained.

Of the several combinations of two validating criteria, that of the classroom observations and pupils' attitudes correlates highest (.576) with the Inventory. When compared with each other, however, these two criteria show the least agreement (r equals .326). In closest agreement are the classroom observations and the principals' ratings (r equals .480). The fact that the ratings of the principals and those of the writer agree more closely with each other than either of these agrees with the reactions of the pupils is of interest. Such findings suggest possible differences in standards of evaluation by which adults and children judge the behavior of teachers. Support is also given the view that pupil rating is an important and unique contribution to an adequate evaluation of teacher behavior.

Table 12 also presents the correlation coefficients computed when using the Inventory scores obtained by use of the simplified scoring procedure. It is interesting to note that, in each case except one (comparison between Inventory and classroom observations), these correlations are slightly higher than those obtained when using the original Inventory scores. These slight differences, though quite consistent, are inconsequential, however, in comparison with the close agreement between the two sets of values.

5. *An Additional Check on the Validity of the Items*

An additional check was made on the validity of the items in the revised instrument. Out of the distribution of the 100 teachers in the unselected group the item response patterns of the twenty with the highest scores and the response patterns of the twenty with the lowest scores were compared, respectively, with the patterns of item response characterizing the original groups of superior and inferior teachers. Chi-square was the statistic used as the measure of comparison.

It seemed reasonable to suppose that, if a particular item were valid, it would call forth similar patterns of response from the two superior groups of teachers and similar patterns from the two inferior groups. Of the 328 (2 x 164) comparisons, 270, or 82 per cent of them had chi-square values less than 9.488, a value which is at the 5 per cent level of significance. A chi-square value less than 9.488, therefore, indicated a high probability that differences found between the two groups, in response to a particular item, could be chance differences. Although a higher proportion than 82 per cent would have been desirable, this result does lend further support to the validity of the items in the Inventory.

TABLE 13

DISTRIBUTION OF RESPONSES OF THE SUPERIOR AND INFERIOR GROUPS OF TEACHERS AND OF THE HIGHEST 20 AND LOWEST 20 TEACHERS IN THE UNSELECTED GROUP TO ITEM 11

11. Pupils are naturally lazy when it comes to doing school work.

	Strongly Agree	Agree	Uncertain	Disagree	Strongly Disagree
Group	*1*	*2*	*3*	*4*	*5*
Superior	0	7	11	73	9 ($N = 100$)
Inferior	3	17	14	63	3 ($N = 100$)
Highest 20	1	2	0	9	8 ($N = 20$)
Lowest 20	1	5	2	10	2 ($N = 20$)

It is of interest to analyzse further, by inspection, the response patterns of an item that seems to be representative of those cases where the chi-square value was greater than 9.488. The response patterns for Item 11 are indicated in Table 13. The chi-square value of 20.618 was found, for Item 11, on comparing the response patterns of the 100 superior teachers with those of the highest twenty in the unselected group. Contributing to the high value of chi-square are the differences between the two groups relative to responses "4" and "5." The great majority of both groups of superior teachers disagree with the statement. Yet, a much larger proportion (40 per cent) of the twenty teachers "strongly disagree" than of the 100 teachers, where only 9 per cent "strongly disagree." This relatively high proportion (40 per cent) undoubtedly accounts for the small representation found under responses "3" and "4" of the group of twenty teachers. As is true with the other items showing high chi-square values, the response patterns indicate validity of the item relative to general agreement or disagreement but reveal discrepancies in the *degree* of reaction. The group of twenty teachers, at each extreme, generally tended to exhibit a more intensive attitude relative to the opinion statements.

B. DETERMINING THE RELIABILITY OF THE INVENTORY

The split-half method was used in determining the reliability of the *Teacher-Pupil Inventory*. Correlating the scores on the odd-numbered items with those on the even-numbered items, a correlation coefficient of .833 was obtained. Use of the Spearman-Brown prophecy formula resulted in a reliability coefficient of .909 for the entire instrument.

The reliability of the Inventory was also determined, by means of the split-half method, on the basis of the scores obtained by the simplified scoring procedure. Under these conditions the odd-even correlation was .794. Determined by use of the Spearman-Brown formula, the reliability coefficient for the entire instrument was found to be .885. The standard error of measurement, determined on the basis of the simplified scoring system, was 6.25. May it be mentioned again at this point that the simplified scoring procedure yielded results which correlated .973 with the original scores.

IV. SUMMARY AND CONCLUSIONS

1. The major purpose of this investigation was to construct a measuring instrument which would gauge the attitudes of teachers toward pupils and serve to differentiate those teachers who get along well with children from those who do not. The instrument, known as the *Teacher-Pupil Inventory,*[1] in its revised form, consists of 164 statements of opinion relating to pupils and children and the part the teacher should play in their educational development. Administration time is from twenty to thirty minutes.

For purposes of standardization and scoring, the original instrument, consisting of 756 items (2 forms with 378 items each), was administered to 100 teachers in each of two groups, one at each extreme of the teacher-pupil relationship continuum, and comparison then made of the item responses. The statistic, chi-square, served as the objective measure of comparison. Those items not differentiating adequately were rejected. A system of scoring was developed on the basis of the responses of the superior and inferior groups of teachers.

2. Significant aspects of the present investigation included the use of empirical procedures of test construction and the use of pupil reaction as a basic validating criterion. The validity of individual items in the instrument was established by the use of the chi-square technique in the comparison of the response patterns provided by the two groups of teachers at the extremes of teacher-pupil relationship. The validity of the Inventory as a whole was determined by administering it to a new sample of 100 teachers from the fourth-fifth-, and sixth-grade levels, *unselected* as to teacher-pupil relationship, on whom ratings were then obtained from each of three sources: (*a*) the principal, (*b*) the writer (from classroom observation), and (*c*) the pupils.

When correlated with Inventory scores, the ratings of the principal, the writer, and the pupils gave validity coefficients of .434, .486, and .452, respectively. After combining the three ratings, the combined result correlated .594 (.596 with simplified scoring) with the Inventory scores. A multiple correlation of .595 between the Inventory and the three criteria was also obtained. These correlations were all significant at the 1 per cent level.

The means of the Inventory scores of the superior, unselected, and inferior groups of teachers were, respectively, 131.0, 77.6, and −32.0. The difference between the means of the superior and inferior groups yielded a critical ratio of 10.8. By statistical test (use of g_1 and g_2), none of the three distributions of Inventory scores was found to depart significantly from normality.

3. The reliability of the *Teacher-Pupil Inventory* was found to be .909 as determined by means of the split-half method and the Spearman-Brown prophecy formula. On the basis of scores obtained by the simplified scoring system, the reliability coefficient was .885.

This investigation has shown that teachers' attitudes toward pupils and their behavior are related to teacher-pupil rapport in the classroom. It has shown further that these attitudes can be measured with as high a validity (.595) as can academic aptitude.

---

[1] The two original forms and the revised form of the *Teacher-Pupil Inventory* may be found in Appendices E, F, and G, respectively, in the writer's Ph.D. dissertation (10) on file in the University of Minnesota Library. Plans are being made to publish the Inventory in the near future.

These findings are of significance in teacher training, selection, and evaluation. It is true that the validity of the Inventory was determined under favorable conditions of test administration. Further research[2] is needed to demonstrate the validity of the instrument for purposes of selection and evaluation of teachers under actual "operating conditions" of administration.

The day may not be far distant when teacher-training institutions will be able to take account of a teacher's personal effectiveness in the classroom in as objective a manner as is done at present with teaching techniques and knowledge of subject matter. It is hoped that such an instrument as the *Teacher-Pupil Inventory* will contribute, in a small but significant way, to this end.

---

[2] Research is under way which should lead to a more refined instrument and a greater knowledge of its measurement possibilities.

BIBLIOGRAPHY

1. BAXTER, BERNICE, *Teacher-pupil relationships.* New York: Macmillan, 1942.
2. BRYAN, R. C. Why student reactions to teachers should be evaluated. *Educ. Adm. & Superv.*, 1941, **17**, 590-603.
3. DROBA, D. D. The nature of attitude. *J. soc. Psychol.*, 1933, **4**, 444-463.
4. DUNLAP, J. W., & DROLL, A. Observations on the methodology in attitude scales. *J. soc. Psychol.*, 1939, **10**, 475-487.
5. FENTON, NORMAN. *Mental Hygiene in School Practice.* Stanford, Calif.: Stanford Univ. Press, 1943.
6. FERGUSON, L. W. The requirements of an adequate attitude scale. *Psychol. Bull.*, 1939, **36**, 665-673.
7. GARRETT, HENRY E. *Statistics in Psychology and Education.* New York: Longmans, Green, 1939.
8. GUILFORD, J. P. *Fundamental Statistics in Psychology and Education.* New York: McGraw-Hill, 1942.
9. KELLEY, IDA, & PERKINS, KEITH. An investigation of teachers' knowledge of and attitudes toward child and adolescent behavior in everyday school situations. *Further Studies in Attitudes,* Series IV, *Stud. in Higher Educ.*, **42**, Purdue University, 1941.
10. LEEDS, CARROLL H. *The Construction and Differential Value of a Scale for Determining Teacher-Pupil Attitudes.* Unpublished Ph.D. thesis on file in the University of Minnesota Library, 1946.
11. LIKERT, R. A technique for the measurement of attitudes. *Arch. Psychol.*, 1932, **22**, No. 140.
12. PRESCOTT, D. A. *Emotion and the educative process.* Washington, D.C.: American Council on Education, 1938.
13. REMMERS, H. H. Generalized attitude scales. *J. soc. Psychol.*, 1934, **5**, 298-312.
14. ROSANDER, A. C. Attitude scale based upon behavior situations. *J. soc. Psychol.*, 1937, **8**, 3-15.
15. SNEDECOR, GEORGE W. *Statistical methods.* Ames, Iowa: Collegiate Press, 1938.
16. STOGDILL, R. M. *Measurement of attitudes toward children.* Ph.D. thesis. Columbus, Ohio: Ohio State Univ., 1934.
17. STRONG, E. K., JR. *Vocational Interests of Men and Women.* Stanford, Calif.: Stanford Univ. Press, 1943. Chapters 23 and 24.
18. SYMONDS, P. M. *Diagnosing personality and conduct.* New York: Appleton-Century, 1931. Chapters 3 and 6.
19. TENENBAUM, S. Test to measure a child's attitude toward school teachers and classmates, *Educ. Adm. & Superv.*, 1940, **26**, 176-188.
20. THURSTONE, L. L., & CHAVE, E. J. *The measurement of attitude.* Chicago: Univ. of Chicago Press, 1929.
21. TIEDEMAN, S. C. Study of pupil-teacher relationship. *J. educ. Res.*, 1942, **35**, 657-664.
22. TSCHECHTELIN, M. A. & OTHERS. Measuring the attitudes of elementary school children toward their teachers. *J. educ. Psychol.*, 1940, **31**, 195-203.
23. WICKMAN, E. K. *Children's behavior and teachers' attitudes.* New York: Commonwealth Fund, 1928.

VOLUME 64
NUMBER 7

WHOLE NO. 313
1950

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

The Nature and Efficacy of Methods of Attack on Reasoning Problems

By
BENJAMIN BURACK
Roosevelt College
Chicago, Illinois

Based on a dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy
at Northwestern University in 1940

Accepted for publication, November 3, 1949

Price $1.00

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

COPYRIGHT, 1950, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION

ACKNOWLEDGMENTS

To Prof. Robert H. Seashore, Chairman of the thesis-committee, for the guidance throughout the whole of this research.

To Professors E. L. Clark, G. L. Freeman, and D. T. Howard, members of the thesis-committee, for suggestions and advice.

To Prof. L. L. Thurstone, for permission to use his unpublished tests of deductive and inductive reasoning ability and for the scoring-keys which he so kindly provided.

To Northwestern University and the Department of Psychology for the Research Fellowship which made it possible to devote full-time attention to this research.

Benjamin Burack

TABLE OF CONTENTS

I. Statement of the Problem 1

II. The Problem-Materials 3

  A. The Induction Problem 3

  B. The Deduction Problem 4

  C. The Geometrical-Figure Problem 5

III. Results 7

  A. The Induction Problem 7

    1. The Methods of Attack Used 7

    2. The Efficacy of the Methods Used 10

  B. The Deduction Problem 10

    1. The Methods of Attack Used 10

    2. The Efficacy of the Methods Used 12

  C. The Geometrical-Figure Problem 13

    1. The Methods of Attack Used 13

    2. The Efficacy of the Methods Used 17

IV. Discussion 19

V. Summary and Conclusions 24

Bibliography 26

I. STATEMENT OF THE PROBLEM*

THE LITERATURE on reasoning reveals a variety of methods used or usable by subjects during problem-solving. Although some of the methods are reported as having been experimentally observed in psychological studies, other methods appear to be personal reflections on how problems could be solved. Whereas the former could actually be either efficacious, indifferent, or detrimental to successful solution, the latter are implicitly presumed to be advantageous. Nine of these "methods of attack" have been selected for investigation in this experiment, in an attempt to determine their use and efficacy for a variety of problems.

There are three questions with which we shall deal:

1. How do individuals differ in the methods they use to solve problems?

2. To what extent are these methods used in solving problems?

3. To what extent are different methods efficacious?

The nine methods selected for investigation are described below.

1. *Clear formulation of the problem.* Dewey early emphasized the importance of awareness of the specific problem. In his first edition of *How We Think,* he discussed as the first 2 of his 5 steps in reflective thinking: (*a*) a felt difficulty, and (*b*) its location and definition.

2. *Preliminary survey of all aspects of the presented material.* The Gestalt School has stressed the role of survey of the whole field—the situation as a whole. And Durkin (6) found that solution was facilitated by initial familiarity with all the parts of her construction-puzzles.

3. *Analysis into major variables.* Seashore (17, p. 554) writes, "In any large and difficult problem, one of the first steps is to analyze out its main variables, which may then be attacked one at a time, in simpler fashion." Here the problem-solver might ask himself this question: "What are the principal factors which determine or influence the solution of this problem?" It is recognized that the problem-solver may utilize variables not helpful for solution.

4. *Locating a crucial aspect of the problem:* "Loci" (Ruger), "Direction" (Maier), "Analysis" (Heidbreder). The subject might question: "What part of the problem seems to merit special attention as the central factor relative to solution?" Ruger found his subjects becoming aware of the "locus" of difficulty. Maier demonstrated the importance of the "way the problem is attacked." However, Maier also warns against "habitual behavior" or "persistence" as *false direction,* wrong attack, a side-track away from correct direction leading to solution. And Heidbreder suggested that the function of analysis is to reduce the supply of stimulus-content by singling out possibly-significant features. As with major variables, the problem-solver may pursue aspects "crucial" to him, but actually not helpful to solution.

5. *Application of past experience.* The significance of applying past experience ("transfer") is a commonplace (Barlow, Chant, Kinnamon, Lindworsky, Lyon). Drever (5) concluded that the application is not necessarily conscious.

6. *Varied trials.* By definition, repeti-

---

* In accordance with new editorial policy, the review of the literature on reasoning, forming the background of this inquiry, is omitted in this monograph, but is a part of the original dissertation, available on interlibrary loan from the Northwestern University Library, Evanston, Illinois.

tion of the same kind of trial could not be successful unless it was correct the first time (also note the previously mentioned warning of Maier against persistence in an unsuccessful direction). Variability has been established as one of the prime factors conducive to solution (Bain, Durkin, Lloyd Morgan, Thorndike, J. C. Peterson, *et al.*).

7. *Control.* In experimentation, control (variation, isolation, repetition) is foundational. In solving a problem, the question can be put: "How can I control each factor involved so as to determine its effect and relative influence?"

8. *Elimination of sources of error.* The literature reveals scarce mention of the role of this factor. What sources of error (for instance, unsteadiness in a delicate mechanical puzzle) may be eliminated or controlled so as to facilitate solution of the problem? As with other methods (see 3 and 4, above), the problem-solver may eliminate factors which are not actually sources of error.

9. *Visualization.* Chen, Comstock, Eidens, Pear, Storring, Wilhelm, and many others have demonstrated the role of imagery in reflective thinking. In behavioristic terms imagery may be considered as the use of implicit trials.

Even at the outset, it appeared that some overlapping exists among certain of these methods, for example "analysis into major variables" and "locating a crucial aspect of the problem."

The subjects were 25 students enrolled in an advanced undergraduate course in psychology. It was recognized that any conclusions concerning efficacy, and especially extent of use, of specific methods would be based on a highly select group of subjects, and that the experiment should be considered only as an exploratory study opening the way to systematic study of problem-solving in the general population.

Three problems were chosen, not as representative of typical human problems, but because their simplicity would permit some objective analysis relatively uncomplicated by the non-intellectual (e.g., emotional) factors often affecting daily problem-solving. The three problems, involving induction, deduction, and geometrical-figure analysis, are described in the next section.

Our problem, then, was to determine the use and efficacy of nine known "methods of attack" on different problems.

II. THE PROBLEM-MATERIALS

A. The Induction Problem

1. *Construction*

THE INDUCTIVE PROBLEM, largely concept-formation (devised by Professor Thurstone), is presented in full, below. Five groups of letters are presented, with four groups always containing something in common; the subject is to *detect the group not containing the feature common to the other four groups.* Exclusive of the practice-problems, 19 such problems were given.

Letter-Grouping

In the five groups of letters below, notice that four of the groups have something in common. The one which is not like the other has been marked.

YLNPJ	XBPZR	YNTFH	YRGSW	YHDLY

Four of the groups begin with the letter Y. The one which does not begin with Y is marked. Look at the next problem.

AAXVP	FNTBB	HQTQX	JDSWR	MECTE

Four of the groups have a repeated letter. The first, second, third, and fifth groups have the letters AA, BB, QQ, and EE respectively. The fourth group does not have a repeated letter. It is unlike the other groups so it is marked.

Work the next two problems. Put a mark under the group which is different in each case.

ABCDE	XYZPS	EFGHI	CXVTN	PJKLM
EDCBA	VWXYZ	IHGFE	JKLMN	SBTGK

In the first problem four of the groups have letters occurring in alphabetical order. You should have put a mark under the fourth group.

In the second problem four of the groups are in alphabetical order, left to right or right to left. The fifth group should be marked.

Look at the next two problems. Try to find which one should be marked.

BDHQX	SDAPZ	GXMOT	LEZTQ	NXUFS
KCUPE	FZEGO	UMBXA	APIQT	LNUKH

In the first problem four of the groups each contain a vowel. The first group should be marked.
In the second problem four of the groups have two vowels. Group five should have been marked.
Now work the next two problems. In each case put a mark under the group which is different.

JPIDX	KPHMA	LPZSU	MPLJE	NPRWO
PZDDD	SSSPB	JPPPN	HNNNX	BQBFB

In the first problem four of the groups have the vowel in the last place. The first group should be marked.

In each group of the second problem there is a letter occurring three times. In four of the groups the three identical letters are together. The last group should be marked.

Work the problems which follow.

1. VBCDF	WQRSH	XIJKF	YLMNF	ZABCF
2. ELMNE	FABCF	GOPQG	HZTMH	IVWXI
3. LGBGM	TBKAK	ZDNEN	FRLLC	VPSPW
4. XVTQZ	BFDHJ	HDFBJ	JBFDH	HJFBD
5. BCAHV	BCOHV	DZELM	DKILM	DTCLM
6. FCRCT	CJCCL	CPCSC	CCRCQ	CKCFC
7. BWBRB	BBCBS	RBMBB	BNBBV	XBBZB
8. BQHTM	AUIEO	ZDKRN	TMPZB	HXKQN
9. EFGHI	EDCBA	UTSRQ	ALRXI	OPQRS
10. CEDGH	FAGHB	DKCES	LMCED	MPCEK
11. CFAGG	PESRP	VROVF	NOMSN	PROTS
12. ONMLK	LKJIH	UTSRQ	GKMOQ	HGFED
13. ALMNA	PBCDP	EVRHE	UJKLU	IDVLI
14. CCPPR	MSSMR	TRRTR	JNNJR	WWDDR
15. BSQVM	TDPDX	GHGRG	FPSFR	MPMBM
16. ACTRM	ABSPQ	CDTGH	PQBXY	MNCST
17. JKLMV	BPQRS	KNWXY	BCDQT	CXTRF
18. FGHLM	CDFGH	JKMNO	STVWX	PQSTU
19. VMSAT	DGMSB	SPLOM	QTMSC	MSDHP

2. *Administration*

This group of problems was administered individually to 20 of the 25 subjects (5 students reported familiarity with this induction-test; their results were excluded). As per directions on the form, the subject first worked all of the practice-problems. Starting with the last set of (five) practice-problems, the subject was questioned, as soon as he finished each one, as to why he had selected the one group of letters which he had marked. His answer was recorded as either "correct," or in full if he had marked the right group but gave an incorrect reason or if he had marked the wrong group and gave some reason for so doing. After he was finished, the following four questions were asked so as to detect evidence of use of methods of attack:

1. What are you supposed to do in these problems? (To detect "Clear formulation of the problem.")

2. Try to list all the ways you can remember in which four groups had something in common lacking in the fifth group. (To detect "Analysis into major variables.")

3. Have you ever had any previous experience with a similar kind of problem? (If yes:) Were you able to apply this past experience here? How? (To detect "Application of past experience.")

4. Do you remember anything else concerning the procedure you used which you believe was helpful in solving these problems? (To detect any other methods of attack.)

Since the three specific questions (exclusive of the last general question) tested for only three methods, we relied upon careful examination of the subject's worksheet to detect evidence of other methods used. It was also recognized that each question was only *one* way of testing for use of any particular method of attack, that subjects unable to state use of a method may use the method, perhaps unconsciously, and that other questions and ways of detecting a specific method could be devised; these limitations are discussed below, under "Discussion" and "Summary and Conclusions." As with the other problems administered, the subject was allowed as much time as he wished for these induction-problems.

B. THE DEDUCTION PROBLEM

1. *Construction*

The deductive problem (selected as one of 15 items in one of Professor Thurstone's Deductive Reasoning Tests) was presented in printed form, as follows:

(*a*) Captain Watts and his son James have been found shot—the father in the chest and the son in the back. (*b*) Both clearly died instantaneously. (*c*) A gun fired close to the person—as, for example, when a man shoots himself—will blacken and even burn the skin or clothes; fired from a greater distance it will leave no such mark. (*d*) The two bodies were found near the middle of a large hall used as a rifle range. (*e*) Its floor is covered with damp sand which shows every footprint distinctly. (*f*) Inside the room there are two pairs of footprints only. (*g*) A third man standing just outside the door or window could aim at any part of the room, but the pavement outside would show no footmarks. (*h*) Under Captain Watts' body was found a gun; no such weapon was found near James. (*i*) In each case the coat, where the bullet entered, was blackened with gunpowder, and the cloth a little singed. (*j*) Captain Watts was devoted to his son and would have died sooner than harm him purposely; hence it is impossible to suppose that he killed him deliberately, even in self-defense. (*k*) But some think that James secretly disliked his father and hoped to inherit his fortune at his death.

a. To what was Captain Watts' death due? Murder? Accident? Suicide?

b. To what was James' death due? Murder? Accident? Suicide?

2. *Administration*

This problem was administered individually to all 25 students. Then the following questions were asked:

1. After reading the problem-questions at the end of the story, did you re-read the entire paragraph, keeping in mind constantly the three possibilities of *murder, accident, suicide*? (Since it early became evident that all subjects understood the problem set to them here, we used this question as probing for the possibility of extreme fixation of the answer-possibilities, or extremely vivid "Formulation of the problem.")

2. How many times did you read the whole paragraph? (To detect careful "Preliminary survey of all aspects of the presented material.")

3. Did you conclude that certain bits of given information were important or relevant, either directly or indirectly, for solving this problem? If so, consult the paragraph, and list all such points. (To detect "Locating a crucial aspect of the problem.")

4. Have you ever had any previous experience with a similar kind of problem? If so, were you able to apply this past experience here? How? (To detect "Application of past experience.")

5. Did you try to apply the possibility of *each* of the three causes—*murder, accident, suicide*—to the death of Captain Watts—and then to the death of James—to determine whether the cause could have been *murder*, could have been *accident*, could have been *suicide*? (To detect "Varied trials.")

6. If you re-read the paragraph, did you single out *each sentence* to try to determine whether any of the various sentences made impossible one or two of the three possible causes? (To detect "Control.")

7. Consult the paragraph for the following questions:

a. Was there any information in the paragraph which makes the following sentence *not* helpful in solving the deaths: "A third man standing just outside the door or window could aim at any part of the room but the pavement outside would show no footmarks."

b. Was there any information which indicated whether *Captain Watts* or *James* died first?

c. Was there any information which made *murder* improbable for James' death?

d. Was there any information which made *suicide* improbable for James' death? (To detect "Elimination of sources of error.")

8. Did you find yourself visualizing the various persons and events? (To detect "Visualization.")

9. Do you remember anything else concerning the procedure you used? (To detect any other methods of attack.)

C. THE GEOMETRICAL-FIGURE PROBLEM

1. *Construction*

The geometrical-figure problem (suggested by Lindley's (10) study) is presented in Figure 1. As with many other similar figure-problems, this figure is to be drawn without going over any line already drawn and without taking the pencil off the paper. Starting at any point except either of the two inner diagonal points makes solution impossible. Fourteen trials were possible on the sheet.

2. *Administration*

This problem was administered individually to 24 of the 25 students (one subject accidentally started at the correct starting-point; his paper was excluded). The subject was allowed 14 trials. If he solved the puzzle on or before the fourteenth trial, he was asked to continue, and was stopped after making 3 successes in a row. The experimenter kept a record of the starting and finishing points for each trial. Then the following questions were asked:

1. Did you assume at the outset that the directions required that you start at some specific corner? (Since pre-experimental investigation had revealed that many subjects started at the upper left-hand corner, we included this question to detect any invalid "Formulation of the problem.")

2. Before beginning the puzzle, did you make a preliminary survey of all aspects of the pictured puzzle, or did you start drawing after a glance at the puzzle? (To detect "Preliminary survey of all aspects of presented material.")

3. Did you conclude at any time that some special part of the puzzle was best to start at? If so, what part? (To detect "Locating a crucial aspect of the problem.")

4. Have you ever had any previous experience with a similar kind of problem? If so, were you able to apply this past experience here? How? (To detect "Application of past experience.")

5. Did you discover certain moves which made solution impossible, or moves which you decided to avoid? If so, what moves or combination of moves? (To detect "Elimination of sources of error.")

6. Did you ever visualize the various possible next-moves before even tracing out these moves on the paper? If so, about how many moves ahead did you visualize on the average? (To detect "Visualization.")

7. Do you remember anything else concerning the procedure you used which you believe was helpful in solving this puzzle? (To detect any other methods used.)

III. RESULTS

Because it would break continuity to list the findings on extent of use for each of three problems and then have to return to the problems for data on efficacy, we present here the complete results for each of the three problems. In the "Discussion" of this paper, we shall return to our two main questions: *use* and *efficacy* of various methods of attack.

A. The Induction Problem

Table 1 shows the distribution of scores. The range is small probably because the items, selected from a speed-test, were taken in unlimited time and were solved, in most cases, by almost all the subjects.

TABLE 1

Scores	No. of Subjects
19	4
18	6
17	4
16	3
15	1
14	2
	$N=20$

Mean = 17.15
Stand. Dev. = 1.53

1. *The Methods of Attack Used*

As contrasted with the other two problems to be discussed, the induction problem revealed only three distinct methods of attack.

(*a*) *Clear formulation of the problem.* The initial instructions read: "In the five groups of letters below, notice that four of the groups have something in common. The one which is not like the others has been marked.

YLNPJ XBPZR YNTFH YRGSW YHDLY

Four of the groups begin with the letter Y. The one which does not begin with Y is marked."

The other examples given (8 altogether) were similar. For each problem, four of the five groups have something in common lacking in the remaining group. It must be noted that the fifth group must be not merely different, but different in the sense of lacking a characteristic common to the other groups; *each* group was probably different from the other four in the sense of containing, say, some letter absent from the other four groups, but this kind of difference was not the required kind. The failure to make the distinction between merely being different and being different in the sense of *lacking a characteristic common to four groups* became evident in the results. A majority of the subjects formulated the problem simply as: Find the group which is different. We have considered this as an "incomplete" formulation.

Only 40% of the subjects correctly formulated the problem, verbally. However, incomplete *verbal* formulation did not seem to involve *actual* incomplete formulation during the working of the problems, as can be observed by the fact that even the lowest score was: 14 correct out of a possible 19, or 74% correct (chance, for 5 choices, would permit 20% correct); it seems axiomatic that successful solution of this kind of problem would require understanding the task. Apparently the nature of the task was evident in spite of the incomplete verbal formulation. This difficulty of mistaking *inability to state a method* for *failure to detect or use a method* was anticipated (see Administration of the Induction Problem, above), and continued to obscure the use of methods in all three problems, as further described in this paper. Incomplete formulation led, however (in a few cases), to the kind of error which might be expected: selecting a group which was merely different. The three examples found are given below:

SUBJECT RLD: BCAHV BCOHV DZELM DKILM DTCLM

The subject's reason: "The others do not have a letter Z." Note that to carry out his incorrect formulation of the problem at this point the subject would be obliged to mark the first group as alone containing an A, the second group as alone containing an O, the fourth group as alone containing a K, and the last group as alone containing a T. Selecting the third group because it alone contained the letter Z indicates incomplete (if not incorrect) formulation. This subject's verbal formulation was: "Pick the one group that is different from the other four groups."

SUBJECT MN: CCPPR MSSMR TRRTR JNNJR WWDDR

Although the group marked is the correct one, the subject's reason is wrong, close however to borderline correctness: "This group has the same letter three times." The correct reason is: The other four groups each have one and only one letter R, or (another reason): The other four groups each have three different letters. This subject's verbal formulation was: "To find the group which is different."

SUBJECT MEM: BSQVM TDPDX GHGRG FPSFR MPMBM

The subject's reason: "The other four groups do not have any letter repeated three times." As a matter of fact her statement is false, for the third group has the letter G three times. But even if her statement were true, her reason indicates the same reversal of the problem-task: she finds a characteristic absent from four groups when the instructions require her to find a characteristic present in four groups. Her verbal formulation was: "Underline the one that is different in any respect from the other four groups."

(*b*) *Preliminary survey of all aspects of the presented material.* The particular nature of this problem required every subject to survey all 5 groups of letters; this factor was found in 100% of the subjects.

(*c*) *Analysis into major variables.* Our question asked the subject to list all the ways in which four groups had something in common lacking in the fifth group. Analysis of all the problems (8 practice-problems and 19 test-problems) gave the following list as including every possible reason for the 27 problems that each subject encountered:

Four groups have some common letter (or 2, 3, 4, or 5 letters).

Four groups have only one of a letter common to all five groups.

Four groups have 2 given letters adjacent.

Four groups have some common letter in a constant position (say, fourth in the group of letters).

Four groups have some vowel (or 2 vowels).

Four groups have some vowel in a constant position.

Four groups have five consonants.

Four groups have three different letters.

Four groups have a repeated letter (or a letter three times).

Four groups have a repeated letter adjacent (or separated).

Four groups have a repeated letter separated by one letter.

Four groups have an alphabetical sequence (of 2, 3, 4, or 5 letters).

Four groups have 2 alphabetical sequences of 2 and 3 letters, respectively.

Analysis of these thirteen ways in which four groups had something in common further revealed 8 major variables basic to all such problems:

Letters.

Vowels.

Consonants.

Position of a letter.

Alphabetical sequence.

Adjacency or separation of 2 given letters.

Quantity of a given letter.

Quantity of different letters within a group.

We credited a subject for each such variable stated or indicated in his list of "ways in which four groups had something in common."

Each of the 8 variables was acknowledged by one or more subjects. Some subjects were more thorough in formulating a variable; for example, some remembered and stated that alphabetical sequence could be either forward or backward.

The distribution of the number of major variables reported is shown in Table 2.

Although other methods of attack may have been employed by the subjects, no other methods revealed themselves as having been used on this problem. Experience with this problem leads us to suggest the following reasons for the fact that only three methods appeared for this problem:

1. Locating a crucial aspect of the problem: Here, it seems (to the writer) that this particular problem does not contain a "crucial aspect" (further explained in the "Discussion").

2. Application of past experience: The 20 subjects had reported, before the experiment, that they had not previously worked problems of this kind.

TABLE 2

Major Variables Reported	Number of Subjects
7	4
6	7
5	6
4	3
	$N = 20$

Mean $= 5.6$ variables
Stand. Dev. $= .97$

3. Varied trials: Variation could mean several things in this problem. (*a*) If it be considered as referring to the subject's questioning each group's lack of a characteristic common to the other four groups, then the small amount of stimulus-material probably permitted every subject to try all five groups, time being unlimited. (*b*) If variation be considered as referring to trying various generalizations (of characteristics possibly common to four of the five groups), then this "method" becomes synonymous with "Analysis into major variables," reported, above, for this problem.

4. Control: Same as for "Varied trials."

5. Elimination of sources of error: As with (1) above, the writer doubts that this factor is adequately meaningful for this problem.

6. Visualization: The stimulus-material itself is visual in nature.

The full implications of the apparent absence of certain methods on certain problems will be more completely treated in the"Discussion."

The range of the total number of methods used per subject was from two to three, with the mean at 2.4 methods per subject.

A question of special interest was: Why did our subjects make the various errors we found on their work-records? We have analyzed the 19 errors made, in relation to the reasons the subjects gave in each case for marking the group of letters they selected:

	Errors	
Perceptual (visual) errors:	12	(63%)
Wrongly formulated problem-task:	3	(16%)
Miscellaneous (hunches, etc):	4	(21%)
Total:	19	

63% of the errors were perceptual mistakes. We list here some examples:

VBCDF WQRSH XIJKF YLMNF ZABCF

SUBJECT PHK: "Because the other groups have the three middle letters in alphabetical sequence."
But so does the one she marked.

LGBGM TBKAK ZDNEN FRLIC VPSPW

SUBJECT KLJ: "Because the other four groups have one letter repeated."
But so does the one she marked.

FGHLM CDFGH JKMNO STVWX PQSTU

SUBJECT JAH: "Because this group is the only one with a vowel."
But note the last group.

16% (3 errors) were the result of wrongly formulated problem-task and have been already listed and discussed above under "Clear formulation of the problem."

2. *The Efficacy of the Methods Used*

Following are the correlations between the methods used and the number of problems worked correctly.

(*a*) *Clear formulation of the problem.* Clear verbal formulation correlated (biserial correlation) —.10 with problems worked correctly, so individual differences in verbal formulation in this experiment do not seem to be related to successful solution. Apparently subjects understood the task in spite of inability to state the task clearly and correctly.

(*b*) *Preliminary survey of all aspects of presented material.* All subjects used this factor, so no comparative evidence was available concerning its efficacy.

(*c*) *Analysis into major variables.* The number of major variables reported correlated (product-moment correlation) +.51 (P.E. = .11) with problems worked correctly, so detection of major variables seems to be moderately correlated (positively) with successful solution.

B. The Deduction Problem

Following is the distribution of scores on the problem concerning the deaths of Captain Watts and his son James.

	Students	
Correct cause for both deaths:	15	(60%)
Correct cause for one death:	10	(40%)
Correct cause for neither death:	0	

Given unlimited time, the majority, 60%, solved the deaths of both Captain Watts and his son James. Of the minority, 40%, some ascertained the correct cause of Captain Watts' death, and some ascertained the correct cause of James' death.

1. *The Methods of Attack Used*

All except one method (Analysis into major variables) were found.

(*a*) *Clear formulation of the problem.* Apparently the nature of the task was self-evident, for all of the subjects stated the problem-task correctly. Two subjects seemed to have formulated the problem more vividly by extreme fixation of the answer-possibilities: these two subjects reread the entire paragraph while keeping in mind the three possible causes (murder, accident, and suicide).

(*b*) *Preliminary survey of all aspects of the presented material.* Table 3 shows the distribution of the number of times the whole clues-paragraph was read. Thus, about half of the subjects did not read through the entire paragraph even a second time. All of the subjects, however, referred back to individual parts of the paragraph.

(*c*) *Locating a crucial aspect of the*

TABLE 3

No. of Readings	No. of Subjects
4	2
3	3
2	8
1	12
	$N=25$

Mean = 1.8 readings

problem. The question asking the subjects to list each item of information important to solution involved every item except two: sentence (g) concerning a third man, and sentence (k) concerning James' dislike of his father. Of the 9 important points, Table 4 shows the distribution of the number of points reported by the subjects. As an average, the subjects reported more than ⅔ of the important items of information.

TABLE 4

No. of Important Points Reported	No. of Subjects
9	1
8	5
7	10
6	5
5	3
4	0
3	1
	$N=25$

Mean = 6.7 points
Stand. Dev. = 1.26

(*d*) *Application of past experience.* Only 3 subjects reported such application. The remarks are quoted:

SUBJECT MEM: "Yes, like the fact that a man can't shoot himself in the back."

SUBJECT SEC: "Yes, I have built up a habit of considering every statement."

SUBJECT RMS: "Yes, the value of reviewing the facts by reading them again."

(*e*) *Varied trials.* On the question "Did you try to deliberately apply the possibility of *each* of the three causes—*murder, accident, suicide*—to the death of Captain Watts—and then to the death of James—to determine whether the cause could have been *murder,* could have been *accident,* could have been *suicide,*" we found that 28% (7 subjects) answered "Yes" and 72% (18 subjects) answered "No." Thus only about ¼ of the subjects felt that they had tried to consider the possibility of each of the three causes for each man's death.

(*f*) *Control.* On the question[1] "If you re-read the paragraph, did you single out each sentence to try to determine whether any of the various sentences made *impossible* one or two of the three possible causes," we found that 44% (11 subjects) answered "Yes" and 56% (14 subjects) answered "No." Thus a little less than half the subjects reported the use of this particular method of control.

(*g*) *Elimination of sources of error.* The four questions testing for ability to detect and eliminate sources of error involved a total of 8 items in the paragraph. Table 5 shows the distribution of the number of points correctly reported by the subjects. Thus, of a maximum possible of 8 items, the subjects, as an average, detected half the items.

(*h*) *Visualization.* 84% (21 subjects) reported visualization of the persons and events in the story, while 16% (4 subjects) reported absence of visualization.

The general question (concerning use of other methods) brought forth remarks by only three subjects. The first subject, on noticing that she was to solve a "detective story," said that she had a good

---

[1] Here, as in the other questions probing for use of methods, the particular question asked is, of course, only one way of searching for methods, in this case for control of individual clues.

63% of the errors were perceptual mistakes. We list here some examples:

VBCDF WQRSH XIJKF YLMNF ZABCF

SUBJECT PHK: "Because the other groups have the three middle letters in alphabetical sequence."
But so does the one she marked.

LGBGM TBKAK ZDNEN FRLIC VPSPW

SUBJECT KLJ: "Because the other four groups have one letter repeated."
But so does the one she marked.

FGHLM CDFGH JKMNO STVWX PQSTU

SUBJECT JAH: "Because this group is the only one with a vowel."
But note the last group.

16% (3 errors) were the result of wrongly formulated problem-task and have been already listed and discussed above under "Clear formulation of the problem."

2. *The Efficacy of the Methods Used*

Following are the correlations between the methods used and the number of problems worked correctly.

(*a*) *Clear formulation of the problem.* Clear verbal formulation correlated (biserial correlation) −.10 with problems worked correctly, so individual differences in verbal formulation in this experiment do not seem to be related to successful solution. Apparently subjects understood the task in spite of inability to state the task clearly and correctly.

(*b*) *Preliminary survey of all aspects of presented material.* All subjects used this factor, so no comparative evidence was available concerning its efficacy.

(*c*) *Analysis into major variables.* The number of major variables reported correlated (product-moment correlation) +.51 (P.E. = .11) with problems worked correctly, so detection of major variables seems to be moderately correlated (positively) with successful solution.

B. The Deduction Problem

Following is the distribution of scores on the problem concerning the deaths of Captain Watts and his son James.

	Students	
Correct cause for both deaths:	15	(60%)
Correct cause for one death:	10	(40%)
Correct cause for neither death:	0	

Given unlimited time, the majority, 60%, solved the deaths of both Captain Watts and his son James. Of the minority, 40%, some ascertained the correct cause of Captain Watts' death, and some ascertained the correct cause of James' death.

1. *The Methods of Attack Used*

All except one method (Analysis into major variables) were found.

(*a*) *Clear formulation of the problem.* Apparently the nature of the task was self-evident, for all of the subjects stated the problem-task correctly. Two subjects seemed to have formulated the problem more vividly by extreme fixation of the answer-possibilities: these two subjects reread the entire paragraph while keeping in mind the three possible causes (murder, accident, and suicide).

(*b*) *Preliminary survey of all aspects of the presented material.* Table 3 shows the distribution of the number of times the whole clues-paragraph was read. Thus, about half of the subjects did not read through the entire paragraph even a second time. All of the subjects, however, referred back to individual parts of the paragraph.

(*c*) *Locating a crucial aspect of the*

TABLE 3

No. of Readings	No. of Subjects
4	2
3	3
2	8
1	12
	$N=25$

Mean = 1.8 readings

problem. The question asking the subjects to list each item of information important to solution involved every item except two: sentence (g) concerning a third man, and sentence (k) concerning James' dislike of his father. Of the 9 important points, Table 4 shows the distribution of the number of points reported by the subjects. As an average, the subjects reported more than ⅔ of the important items of information.

TABLE 4

No. of Important Points Reported	No. of Subjects
9	1
8	5
7	10
6	5
5	3
4	0
3	1
	$N=25$

Mean = 6.7 points
Stand. Dev. = 1.26

(*d*) *Application of past experience.* Only 3 subjects reported such application. The remarks are quoted:

> SUBJECT MEM: "Yes, like the fact that a man can't shoot himself in the back."
>
> SUBJECT SEC: "Yes, I have built up a habit of considering every statement."
>
> SUBJECT RMS: "Yes, the value of reviewing the facts by reading them again."

(*e*) *Varied trials.* On the question "Did you try to deliberately apply the possibility of *each* of the three causes—*murder, accident, suicide*—to the death of Captain Watts—and then to the death of James—to determine whether the cause could have been *murder,* could have been *accident,* could have been *suicide,*" we found that 28% (7 subjects) answered "Yes" and 72% (18 subjects) answered "No." Thus only about ¼ of the subjects felt that they had tried to consider the possibility of each of the three causes for each man's death.

(*f*) *Control.* On the question[1] "If you re-read the paragraph, did you single out each sentence to try to determine whether any of the various sentences made *impossible* one or two of the three possible causes," we found that 44% (11 subjects) answered "Yes" and 56% (14 subjects) answered "No." Thus a little less than half the subjects reported the use of this particular method of control.

(*g*) *Elimination of sources of error.* The four questions testing for ability to detect and eliminate sources of error involved a total of 8 items in the paragraph. Table 5 shows the distribution of the number of points correctly reported by the subjects. Thus, of a maximum possible of 8 items, the subjects, as an average, detected half the items.

(*h*) *Visualization.* 84% (21 subjects) reported visualization of the persons and events in the story, while 16% (4 subjects) reported absence of visualization.

The general question (concerning use of other methods) brought forth remarks by only three subjects. The first subject, on noticing that she was to solve a "detective story," said that she had a good

---

[1] Here, as in the other questions probing for use of methods, the particular question asked is, of course, only one way of searching for methods, in this case for control of individual clues.

TABLE 5

No. of Points Reported	No. of Subjects
6	3
5	7
4	8
3	5
2	1
1	0
0	1
	$N=25$

Mean = 4.1 points
Stand. Dev. = 1.32

imagination. Her solution would be satisfactory except for ignoring the gunpowder-burns on the back of James' coat: "Captain Watts accidentally shot himself in the chest. The bullet went through his body and hit James, standing behind him, in the back."

The remarks of the other two subjects (both of whom solved the problem correctly) suggest, again, that subjects may not be aware of using methods and may not be aware of how they actually solve a problem: "Just reading and restating the points seemed to suggest solution," and "Reading the material suggested various possible solutions which I checked on."

The only method not discovered was "Analysis into major variables." For this particular problem this method could refer to the three ways in which each man could have died (murder, accident, suicide); then the method would be synonymous with "Varied trials," reported on this problem. On the other hand, if the important clues are considered as major variables, then this method would be synonymous with "Locating a crucial aspect," reported on this problem.

The range of the total number of methods used per subject was from 3 to 7, with the mean at 4.7 methods per subject. Of course this average conceals the qualitative differences in the use of a given method by different subjects and also the quantitative difference in the extent to which a method was used.

2. *The Efficacy of the Methods Used*

The following tables and bi-serial correlations show the relationship between the use of a method and the success or failure in solving the two deaths.

(*a*) *Clear formulation of the problem.* All subjects seemed to formulate the problem clearly, so no measure of efficacy was possible. It might be noted that the two subjects who formulated the problem quite vividly (by fixating the three answer-possibilities while re-reading the paragraph) both solved the problem.

(*b*) *Preliminary survey of all aspects of the presented material.* The number of times the subject read the clues-paragraph correlated +.22 with successful solution, so complete re-survey seems slightly related to successful solution.

(*c*) *Locating a crucial aspect of the problem.* The number of important points detected and reported correlated +.39 (P.E. = .15) with successful solution, so detection of this kind of crucial aspect seems slightly related to successful solution.

(*d*) *Application of past experience.* Only 3 subjects reported the direct application of past experience (2 solved the problem), so no measure of the efficacy of past experience was feasible.

(*e*) *Varied trials.* In interpreting the following percentages (in Tables 6, 7, and 8), it must be noted that 60% of the entire group solved the problem. Therefore the percentage of users of the method "Varied trials" must be compared

TABLE 6

	Solved (by 60% of the subjects)	Failed (by 40% of the subjects)
7 subjects who tested three possible causes	5 (71%)	2 (29%)
18 subjects not testing three possible causes	10 (56%)	8 (44%)

with *60%*, not with 50% (as would be the case if equal numbers solved and failed the problem). Of the 7 subjects who tested each of the three possible causes, 71% solved the problem, 11% more than might be expected, so (our measure of) variation seems slightly related to successful solution.

TABLE 7

	Solved (by 60% of the subjects)	Failed (by 40% of the subjects)
11 subjects who tested the effect of isolated items	9 (82%)	2 (18%)
14 subjects not testing the effect of isolated items	6 (43%)	8 (57%)

(*f*) *Control.* Of the 11 subjects who isolated items and tested their effects, 82% solved the problem, 22% more than might be expected, so this measure of control seems somewhat related to successful solution.

(*g*) *Elimination of sources of error.* The number of items detected concerning sources of error correlated +.60 (P.E. = .12) with successful solution, so our measure of detection and elimination of sources of error seems to be moderately and positively correlated with successful solution.

(*h*) *Visualization.* Of the 21 subjects who reported visualization of the persons and events in the story, 57% solved the problem, 3% less than might be expected, so visualization does not seem related to successful solution.

TABLE 8

	Solved (by 60% of the subjects)	Failed (by 40% of the subjects)
21 subjects who reported visualization	12 (57%)	9 (43%)
4 subjects not reporting visualization	3 (75%)	1 (25%)

The total number of methods used correlated +.48 (P.E. =.14) with successful solution, and thus was moderately and positively correlated with success in this problem.

Thus, four methods seem somewhat related to successful solution of this particular problem:

Locating a crucial aspect
Varied trials
Control
Elimination of sources of error

C. The Geometrical-Figure Problem

Table 9 shows the distribution of the number of trials required to solve the problem, and includes the 6 subjects who failed to solve the problem in the 14 trials allowed.

1. *The Methods of Attack Used*

This tracing-puzzle is so constructed that, by starting at either one of the two end-points of the inner diagonal, solution is easy. But starting at any other of the 14 remaining points makes solution impossible. Of course it is possible that a subject will happen to start at one of the two correct starting-points; one subject did start thus and was surprised at his early solution, and his paper was excluded from our results since no infor-

TABLE 9

Trials	No. of Subjects
3	4
4	4
5	1
6	2
7	1
9	1
12	1
13	2
14	2
Failed	6
	$N=24$

Median = 8 trials

mation concerning work-methods was available. Of the nine methods of attack, only "Application of past experience" was not reported.

(*a*) *Clear formulation of the problem.* All of the subjects seemed to have formulated the problem clearly; apparently the task in this puzzle is self-evident, from the nature of the instructions employed. Although many subjects started (and tended to continue starting) at the upper left-hand point, they had not assumed that the directions required that they start at this corner; possibly the deeply ingrained reading-habit of starting at upper-left may underlie this phenomenon (Lindley, 10). One special aspect of the problem-task, however, was found to be unclearly formulated by a few subjects, who asked whether they might erase their last line drawn; the directions specified ". . . do not erase or change what you have drawn. . ."

(*b*) *Preliminary survey of all aspects of the presented material.* 38% of the subjects reported that they had studied the puzzle before beginning to draw. In one case a subject pondered for half a minute before beginning the first trial. On this particular problem, the time spent in preliminary survey was very short; subjects tended to gain familiarity with the problem through overt trial-and-error rather than through close inspection or implicit (imagined) trial-and-error.

(*c*) *Analysis into major variables.* Since starting at either of the two correct starting-points makes solution rather easy, there is only one major variable: correct starting-point. However, since it is possible to go into a blind-alley even after starting correctly, the avoidance of blind-alleys may be considered as a minor variable. The subjects reporting the major variable all achieved the same general idea: starting at the center ("it seemed best to start at the center and work around"). However, the minor variable was not so specifically formulated. Although most of the subjects who reported it simply warned against blind-alleys, others were more specific

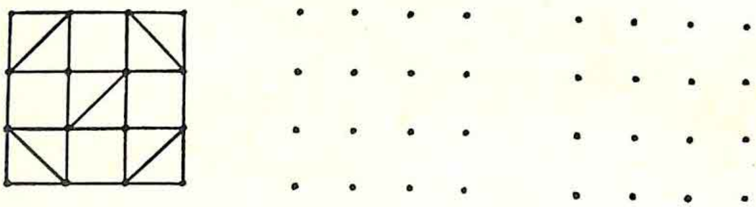


FIGURE 1. Geometrical-figure problem.

Draw this figure, without going over any line already drawn and without taking the pencil off the paper. Fill in as many lines as you can. If you are unable to complete the figure, *do not erase or change* what you have drawn, but start over again on a new figure. *Continue* until you draw a complete figure successfully.

("Leave an open space near a partially drawn section to get back in later to complete the section." And another: "Don't start at any of the 4 outside corners."). 17% of the subjects reported the major variable. 38% reported the minor variable (blind-alleys). And none reported both.

(d) Locating a crucial aspect of the problem. Some subjects hit upon a *"false crucial aspect";* they selected one point for starting and felt that solution depended only upon the care exercised after starting. Another false "crucial aspect" or "direction" consisted of detecting patterns: These subjects noticed the large octagon-pattern and tried to make it first, which necessitated starting at a point which was (unknown to the subject) detrimental to solution. Finally, some subjects conceived the idea of starting in the center, "to work around"; since two of the four center-points were correct starting-points, this was a partial localization of the problem, and the subjects soon solved the puzzle. 50% of the subjects reported some "crucial" aspect (including false ones).

(e) Application of past experience. Although there were a few subjects who had worked a similar kind of puzzle in the past, none reported the application of any specific principle carried over from previous related experiences. Again, however, the criterion (verbalization) may have obscured minor aspects of carryover. Theoretically, the one great application that might have been made would be the notion of *trying different starting points.* Inspection reveals, however, that many "similar" puzzles are so constructed that, *by starting at any point and exercising care,* one can solve the puzzle; if this is true of the geometrical-figure puzzles previously tried by the subjects, then no positive carryover of the major variable (correct starting point) in our puzzle could be expected. On the contrary, we should expect a disregard of the starting-point as a negative transfer to this problem. As reported below, under "Varied trials" and "Control," many of the subjects started consecutive trials from the same point, thus confirming the possibility of negative transfer from previous puzzles in which starting-point may have been unimportant. This finding also suggests that the efficacy of methods of attack may not be general but conditional, depending upon the degree to which any given method makes clear the crucial major variable or variables basic to the problem at hand. A similar finding is reported under efficacy of the factors of "Varied trials" and "Control."

(f) Varied trials, (g) Control. As experimentally defined, "control" ("variation, isolation, repetition") has been broken down into two factors for this problem: (f) Varied trials ("variation") and (g) Control ("isolation, repetition").

Our only expedient measure of "Varied trials" was: the number of different starting-points used, divided by the number of trials required for solution. Thus, the highest possible variability would be 100% for a subject who varied his starting-point on each new trial. The lowest possible variability would be 7% (really representing no variability), for a subject who tried only one starting-point in 14 trials ($1 \div 14$). It should be noted, however, that this measure of variability refers to *variability of starting-point* and that there remains the complex variability displayed *after* starting. This latter variability, important no doubt to the subject, did not permit measurement; thus we were unable to establish a general measure of variability.

Our measure of "Control" was: Total number of trials in which the same starting-point was used two or more times consecutively, divided by the number of trials. Thus, a subject who started at one point for three trials, then at another point for two consecutive trials, and then at another point for, say, five successive trials, has demonstrated 100% control (3 + 2 + 5, divided by 10), in the sense that each new starting-point has been tried out for 2 or more trials before being abandoned. A subject who varied his starting-point on each new trial would demonstrate zero control (and, it should be noted, a variability of 100%). Thus, by definition for our limited means of measurement, high variability tended to go with low control (for this problem). However, since variation is an important aspect of control in the latter's full meaning and use, we have here considered our two measures as two aspects of the same principal factor.

The results showed "Variability" ranging from 14% (only 2 starting-points in 14 trials) to 100% (different starting-point on each trial), with the mean at 56%. "Control" ranged from 0% (changed starting-point on each trial) to 100% (each new starting-point tried for 2 or more consecutive trials), with the mean at 51%. Two interesting examples of introspections are quoted, as indicating personal differences among subjects in carrying out empirically the sensed awareness of the efficacy of variation:

SUBJECT ALG (Variability 36%; Control 43%): On starting her sixth trial at a new point, said, "Variety is the spice of life, so I'll start here." On starting her last possible trial (14th), she said, "Maybe if I start in the middle, something might happen." She started at a correct point and solved the puzzle.

SUBJECT RMS (Variability 14%; Control 93%): This subject began her second trial at a different point, but continued starting at this point right through the next 13 trials, unsuccessfully as regards the solution. While working on her seventh trial, she said, "Maybe if I start at a different place, I'd make it." But she started her eighth trial at the same point again. On starting her tenth trial at still the same point, she said again, "Maybe I don't start at the right place." But she began her eleventh trial at the same point again, as with the twelfth, thirteenth, and fourteenth (last).

Thus, for our measures in this problem "Control" seems to measure mainly *persistence,* with most of the variability partialled out and measured as "Varied trials."

(h) Elimination of sources of error. 42% of the subjects reported the identification and elimination of one source of error, mainly: blind-alleys (the minor variable). An unusual source of error was encountered, and met, by one subject who on her second trial unknowingly omitted the center diagonal, finished the rest of the figure, and thought that she had solved the puzzle until reminded (by the experimenter) of her omission. To meet this source of error, before beginning her next trial she marked light crosses in the five places where diagonals had to be drawn; she encountered no difficulty thereafter. This, of course, also represented "Locating a (minor) crucial aspect."

(i) Visualization. The maximum number of possible-next-moves visualized by the subjects ranged from 2 to 7, with the mean at 4.4 moves.

No other methods were revealed in the answers to the general question.

The range of the total number of methods used per subject was from 3 to 8, with the mean at 5.6 methods per subject. Again, we recognize that the mean of 5.6 blankets the fact that a given method differed qualitatively (and also in quantitative degree) from one subject to another on the same problem.

2. *The Efficacy of the Methods Used*

The following tables show the relation between the use of a method and the number of trials involved in solution. The number of trials are presented in a two-fold classification: Superior subjects (below the median number of trials) and Inferior subjects (above the median number of trials). These small numbers of cases derived from a study mainly qualitative in design do not permit rigorous statistical treatment, but inspection of the results grouped into the 2 categories listed reveals general tendencies toward efficacy or non-efficacy of various methods.

For 3 methods where a rank-difference correlation was feasible (although no measures of reliability were obtainable), the correlation is presented.

(a) Clear formulation of the problem. All subjects seemed to formulate the problem clearly, so no measure of efficacy was possible.

(b) Preliminary survey of all aspects of the presented material. Inspection reveals that among the subjects who reported preliminary survey there were about equal numbers of superior and inferior subjects. And among those not reporting preliminary survey, there were also about equal numbers of superior and inferior subjects. A preliminary survey of all aspects of the stimulus-material (in the minor degree to which it appeared at all) did not seem helpful for solution of this particular problem.

TABLE 10

	Number of Trials Involved in Solution	
	12 Superior Subjects (3 to 7 trials)	12 Inferior Subjects (9 trials to "failed")
Subjects making preliminary survey ($N=9$)	5	4
Subjects omitting preliminary survey ($N=15$)	7	8

(c) Analysis into major variables. The majority (3 out of 4) of those reporting the major variable were superior subjects. More inferior subjects reported the minor variable (unimportant in this problem), and about equal numbers of superior and inferior subjects reported neither variable. Awareness of the major variable seems to be related positively to successful solution.

TABLE 11

	Number of Trials Involved in Solution	
	12 Superior Subjects	12 Inferior Subjects
Detected major variable ($N=4$)	3	1
Detected minor variable ($N=9$)	3	6
Detected neither variable ($N=11$)	6	5

(d) Locating a crucial aspect of the problem. Of those who had an hypothesis which they considered crucial, there were more inferior subjects. And of those lacking such an hypothesis there were

TABLE 12

	Number of Trials Involved in Solution	
	12 Superior Subjects	12 Inferior Subjects
Hypothesis ($N=12$)	5	7
No hypothesis ($N=12$)	7	5

more superior subjects. Mere presence of an hypothesis concerning a "crucial" aspect seems, if anything at all, negatively related to successful solution for this particular problem. It is probably a matter, not of having an hypothesis but of *quality* of the hypothesis, since a poor hypothesis is not merely indifferent but actually detrimental to the speed of solution on the geometrical-figure problem.

(e) Application of past experience. No subjects reported the direct application of specific principles which could be tabulated.

(f) Varied trials. Variability correlated +.87 (P.E. = .07) with successful solution, so varied trials seem to be highly related (positively) to successful solution, because our specific measure of variability (changed starting-point) eventually led to the correct starting-point.

(g) Control. Control correlated −.61 (P.E. = .11) with successful solution, so (our measure of) control seems to be moderately related (negatively) to successful solution, because isolation and repetition of same starting-point prevented the subject from trying the only two correct starting-points.

(h) Elimination of sources of error. Of those reporting a source of error, there were more inferior subjects than superior ones. And of those not reporting a source of error, there were more superior subjects than inferior ones. In stating that elimination of a source of error was not related positively to successful solution, it must be remembered that these errors (mainly blind-alleys) could not prevent failure based almost entirely on a single factor, namely: *starting-point,* a peculiarity of this particular problem.

TABLE 13

	Number of Trials Involved in Solution	
	12 Superior Subjects	12 Inferior Subjects
Detection of source of error ($N=10$)	4	6
No detection of source of error ($N=14$)	8	6

(i) Visualization. The number of visualized possible-next-moves correlated +.37 (P.E. = .13) with successful solution, so such visualization seems to be slightly related (positively) to successful solution.

The total number of methods used correlated −.30 (P.E. = .14) with successful solution, and seems slightly related negatively, apparently because some methods hindered solution. So the mere use of more methods does not seem to guarantee solution.

Thus, four methods show some relationship to successful solution of this problem: (*a*) Analysis into major variables, (*b*) Locating a crucial aspect, (*c*) Varied trials, and (*d*) Control. The *nature* of the relation (positive or negative) reveals that detection of the only important variable (correct starting-point), by means of varied trials rather than by persistence at the same starting-point ("Control"), leads to rapid solution of this puzzle. In summary, although the various methods of attack may be helpful in other problems, only "varied trials" and "Analysis into major variables" seem to be efficacious for solution of this problem, because the first leads to discovery of the second, and the second guarantees solution. The very construction of the figure, offering only 2 correct starting-points out of 16 possible starting-points, seems to render other methods useless.

IV. DISCUSSION

In interpreting the results of this experiment, certain limitations and sources of error must be borne in mind.

1. The particular question or technique we used to determine use of a given method of attack was only one way of determining use of that method.

2. The particular question or technique employed by us often required a dichotomous answer, thus permitting only gross quantification of the use of some methods.

3. There are probably other methods of attack besides the ones for which we searched. None was revealed in our investigation.

4. Methods may be used but not reported. And, contrariwise reported use of a method does not guarantee that it was used by the subject. Our experimental technique revealed itself to be too dependent upon introspective report.

5. In interpreting the absence of efficacy of certain methods of attack, it is important to distinguish between the methods reported by untrained individuals and the possibility of application of these methods by trained problem-solvers.

6. Besides the various methods of attack used, other intellectual factors are of course partly basic to successful solution of these problems, factors not appearing as methods but contributing nevertheless to solution.

7. We used a highly select sample of subjects.

8. The three problems which we used, although different in nature, were hardly representative of typical problems.

In the following interpretations of the results of the experiment, Table 14 will be used for the first two questions, and Table 15 for the third. In Table 14 the first 3 columns show the average degree of use (based on a maximum possible use of 100%) of each method, by the group, for each of the three problems; the missing entries indicate that no measure was obtainable for certain methods in certain problems. The fourth column shows the average degree of use for three problems, and the last column shows the average degree of use for only those problems in which a measure of the method was obtainable.

1. How did individuals differ in the methods they used to solve the problems?

An obscuring factor in the presentation of the comparison of individuals with regard to use or non-use of various methods lies in the fact that the individuals differed, not so much in the kind of methods used, but rather in the extent to which each used a particular method. The individual differences in extent of use of each method for each problem have been indicated in the body of this paper; in summary it can be stated that extent of use for any one method for any one problem appeared to be distributed normally, considering the limitations of measurement (gross measurement, dichotomous answers, etc.). If we credit an individual with use of a method, regardless of degree of use, the differences in the number of methods used on each problem only approximate a normal curve; since each of the three distributions represents, not a single variable, but rather a total use of any or all of 9 variables (methods), we would not necessarily expect normal distributions.

a. In the induction problem, where only 3 methods were used, every subject used either 2 or 3 methods, with the mean at 2.4 methods per subject.

b. In the deduction problem, where 8 methods were used, subject used from 3 to 7 methods, with the mean at 4.7 methods per subject.

c. In the geometrical-figure problem, where 8 methods were used, subject used from 3 to 8 methods, with the mean at 5.6 methods per subject.

2. To what extent were the methods used in different problems?

In Table 14, the fourth column of fig-

TABLE 14

AVERAGE EXTENT OF USE OF METHODS
(All figures based on maximum possible or 100% use)

Method	Induction Problem	Deduction Problem	Geometrical Figure	For Three Problems*	For Relevant Problems*
Clear formulation problem	40%	100%	100%	80%	80%
Preliminary survey	100%	45%[b]	38%	61%	61%
Analysis into major variables	70%[a]	Same as "crucial aspect" or "varied trials"	17%	29%	43.5%
Locating a crucial aspect	Not relevant?	74%[c]	50%	41%	62%
Application of past experience	0%	12%	0%	4%	12%
Varied trials	Same as "major variables"?	28%	56%	28%	42%
Control	Same as "major variables"?	44%	51%	32%	47.5%
Elimination of sources of error	Not relevant?	51%[d]	42%	31%	46.5%
Visualization	Not relevant?	84%	15%[e]	33%	49.5%
Total number of methods used on problem	3	8	8		
Mean-number of methods used per subject	2.4	4.7	5.6		
Range of number of methods used by subjects on each problem	2–3	3–7	3–8		

[a] Based on maximum of 8 major variables.
[b] Based on maximum of 4 readings (highest number by any subject).
[c] Based on maximum of 9 crucial aspects.
[d] Based on maximum of 8 sources of error.
[e] Based on maximum of 29 possible visualized next-moves.
* These figures are intended as suggestive only, and have not been corrected for small differences in the number of subjects in each of the three groups.

ures shows that some methods seem hardly to have been used at all. For example, "Application of past experience" seems to be almost completely absent (4%), and another 5 methods show use of only 28% to 33%. Inspection reveals, however, that "Application of past experience" was reported (by one or more subjects) on only one of the three problems; when the figure is based on the problem where the method was used, we see that 12% of the subjects reported its use. Further inspection reveals that the problem was a human relations situation (death of two men), in which past experience would more likely be applicable, as compared with the two problems of a more perceptual nature. Even in these, however, it is the author's impression that finer techniques of investigation might have revealed subtle use of the methods in question.

For the 5 methods showing use of only 28% to 33%, inspection reveals that each method was measurable on only two problems; when the figures are based on the two problems where used, the range becomes 42% to 49.5%.

Why were measures unobtainable for certain methods on certain problems? Two reasons are suggested.

a. For a given problem, certain presumably different methods appeared (to both the experimenter and the subjects) to carry the same meaning. On the induction problem, for example, "Varied trials" and "Control" (two aspects of trial-and-error) consisted of applying generalizations to the groups of letters; this method was measured as "Analysis into major variables." On the deduction problem, "Analysis into major variables" could be considered as looking for clues, measured as "Locating a crucial aspect," or as questioning each of the three possible causes (murder, accident, suicide) in relation to the death of each man, measured as "Varied trials." It matters little whether the subject's behavior is classified under one or the other of two headings; the important thing is that a distinction should be made between "methods" different in terminology only and methods different in actual behavior of the subjects.

b. A similar point is that, on a given problem certain "methods" seem either meaningless or irrelevant. On the induction problem, for example, the experimenter was unaware of any "crucial aspects" among the groups of letters; if the common characteristics are crucial, then this method is identical with "Analysis into major variables." On the same problem the experimenter was unaware of sources of error in the stimulus-material; and "Visualization" seemed to be irrelevant, in the sense that all the material was visual and seemed to require no additional visualization (imagery). The introspective reports confirmed the absence of such "methods."

We conclude, therefore, that although there is wide variation in the extent to which different methods are used, extent of use is partly a function of relevancy of certain methods to certain problems. In Table 14, for example, inspection reveals that only a few methods (3) are used on the induction problem, whereas most of the methods (8) are used on the other two problems; and the average use per subject follows the total number of methods used. Since the individuals do not change, we suggest that the number of methods used in different problem-situations depends upon the kind of problem involved: (*a*) some problems "permit" use of a given method (for example, application of past experience) while other problems do not; (*b*) in certain situations a given "method of attack" becomes identical in meaning with a differently worded method of attack; and *(c)* in certain situations a given "method of attack" is a term rather than a form of human behavior.

5. To what extent were the different methods efficacious?

Table 15 summarizes the degree of efficacy for each method on each problem.

a. Clear formulation of the problem showed no relationship to solution for one problem and was of unknown efficacy for the other two problems. It is suggested that the task may have been quite obvious in our problems, thereby permitting no individual differences in understanding of the requirements of the problem.

b. Preliminary survey of all aspects of the presented material was of slight help in the deduction problem, of unknown efficacy in the induction problem, and of no efficacy in the geometrical-figure problem. It is suggested that success in the latter two problems requires considerable trial-and-error rather than global survey.

c. Analysis into major variables was moderately helpful in the induction problem, slightly helpful in the geometrical-figure problem, and unmeasura-

TABLE 15

Efficacy of Methods of Attack

Correlations indicated when obtainable. Degree of relationship ("Zero" or "slight") indicated where correlations could not be derived. Relationships not obtainable for: (*a*) methods not measurable and (*b*) methods showing no substantial individual differences.

Method	Induction Problem	Deduction Problem	Geometrical Figure
Clear formulation of problem	−.10	Used by all: efficacy not obtainable	Used by all: efficacy not obtainable
Preliminary survey	Used by all: efficacy not obtainable	+.22	Zero
Analysis into major variables	+.51 (P.E. = .11)		Slightly positive
Locating a crucial aspect		+.39 (P.E. = .15)	Slightly negative
Application of past experience	Used by none	Used by only three: efficacy not obtainable	Used by none
Varied trials		Slightly positive	+.87 (P.E. = .07)
Control		Slightly positive	−.61 (P.E. = .11)
Elimination of sources of error		+.60 (P.E. = .12)	Slightly negative
Visualization		Zero	+.37 (P.E. = .13)
Total number of methods used	Number of methods insufficient to permit correlation	+.48 (P.E. = .14)	−.30 (P.E. = .14)

ble in the deduction problem. It is suggested that this method tends to be helpful because it may expose a variable fundamental to solution (along with unimportant variables).

d. Locating a crucial aspect was slightly helpful in the deduction problem, slightly disadvantageous in the geometrical-figure problem, and unmeasurable in the induction problem. It is suggested that the potential efficacy of the search for crucial aspects is contingent upon the time or trials available to the problem-solver: with unlimited time (as in the deduction problem), the method tends to be helpful; with limited time or trials (as in the geometrical-figure problem), the method may obstruct solution, especially if one hypothesis be perseverated (Maier's "false direction").

e. Application of past experience was not reported by our subjects, so no measure of efficacy was obtainable. On the geometrical-figure problem, however, the work-sheets indicated a tendency to start at the same point (usually the upper-left) on consecutive trials, thus making solution impossible. It is suggested that previous practice in tracing figures acts as negative transference, because most tracing-figures can be solved from any starting-point, so long as care is exercised to avoid blind-alleys.

f. Varied trials was slightly helpful

in the deduction problem, paramount in the geometrical-figure problem, and unmeasurable in the induction problem. Recognizing, however, that in the geometrical-figure problem our only measure of variation was variation of starting point, and that correct starting-point is the only important factor for this particular problem, it is suggested that the efficacy of variation is dependent upon its potentiality of exposing the major variable or crucial aspect of a problem.

g. *Control* was slightly helpful in the deduction problem, detrimental in the geometrical-figure problem, and unmeasurable in the induction problem. It is suggested that the efficacy of "isolation and repetition" of variables felt to be important by the subject, or the efficacy of aspects felt to be crucial, is dependent upon the subject's luck or skill in testing the important variable or aspect.

h. *Elimination of sources of error* was moderately helpful in the deduction problem, slightly disadvantageous in the geometrical-figure problem, and unmeasurable in the induction problem. It is suggested that the efficacy of this method is dependent upon its potentiality of exposing significant features of the situation.

i. *Visualization* showed a small positive correlation with successful solution in the geometrical-figure problem, no effect in the deduction problem, and was unmeasurable in the induction problem. Visualization could be of no help on the geometrical-figure problem unless a subject could visualize all 29 moves, because 2 of the 16 starting-points practically guarantee solution, whereas 14 starting-points make solution impossible with or without visualization. The small positive correlation in the geometrical-figure problem may indicate the typical positive correlation of abilities rather than a causal relationship. It is suggested that, as with several other methods already discussed, the efficacy of visualization (and other imagery) is dependent upon the type of problem.

V. SUMMARY AND CONCLUSIONS

The literature on reasoning reveals various methods, techniques and auxiliary aids used by subjects during problem-solving. Nine of such known "methods of attack" were selected for investigation in this experiment.

1. Clear formulation of the problem.
2. Preliminary survey of all aspects of the presented material.
3. Analysis into major variables.
4. Locating a crucial aspect of the problem.
5. Application of past experience.
6. Varied trials.
7. Control.
8. Elimination of sources of error.
9. Visualization.

Three different problems were selected to be given to the subjects serving in this experiment. The induction problem was a short concept-formation test (identification of similarities and differences in letter-grouping series) devised by Professor Thurstone; 19 items were used. The deduction problem, chosen from one of Professor Thurstone's reasoning tests, presented a paragraph of clues to the death of two men; the subject is to ascertain the cause of the death of each man. The general problem was a geometrical-figure problem of a maze-tracing variety. The problems were administered individually to a group of 25 undergraduate students, with an unlimited time-allowance. A number of specific questions were asked of the subjects during and after the work-period, and objective records of the work-sheets were kept, designed to determine the use of the various methods of attack selected for investigation.

Our problem was to determine the nature and efficacy of the methods of attack used on different problems. The three main questions, and the results, were as follows.

1. How did individuals differ in the methods they used to solve the problems?

a. If we credit an individual with use of a method, disregarding *degree* of use, the number of methods used was not distributed normally for either all three problems combined or for any one problem. It appears to us that normal distributions would not necessarily be expected, considering that each distribution represents not a single variable or continuum, but rather a total use of any or all of nine variables (methods).

b. In the induction problem, every subject used either two or three methods, with the mean at 2.4 methods. In the deduction problem, subjects used from 3 to 7 methods, with the mean at 4.7 methods. In the geometrical-figure problem, subjects used from 3 to 8 methods, with the mean at 5.6 methods.

c. Individual differences in degree of use of a particular method on a given problem appeared to be distributed normally, considering the limitations of measurement.

2. To what extent were the methods used in different problems?

There was wide variation in the extent to which different methods were used, certain methods being used on one problem, and other methods on a different problem. Two methods were used on all three problems, six methods were used on two problems, and one method was used on only one of the three problems. Examination of the tasks required in the three problems suggests that, for a specific problem: (a) certain presumably-different methods appeared to carry the same meaning, and (b) certain so-called "methods" appeared to be meaningless or irrelevant.

3. To what extent were the different methods efficacious?

a. For a given problem some methods were efficacious, some were indifferent, and some were detrimental to solution. The latter case seemed to occur whenever a particular method served to prevent or postpone discovery of the variable or aspect fundamental to solution of the problem.

b. Because no method was used on all three problems (except "Clear formulation," used by almost everyone, and thus of unmeasurable efficacy), efficacy of a given method was demonstrable for only one or two of the three problems. For the latter case (two problems), two methods were efficacious on both problems, two methods were efficacious on one problem and "indifferent" (zero correlation) on the other, and three methods were efficacious on one problem and "detrimental" (negative correlation) on the other.

In interpreting the results, however, certain limitations and sources of error were recognized:

1. The particular question or technique used to determine use of a given method was only one way of determining use of that method.
2. The particular question or technique employed by us often required a dichotomous answer, thus permitting only gross quantification of the use of methods.
3. There were probably other methods used which escaped our tools of search.
4. Some of the nine methods may have been used but not reported and not demonstrated in the work-sheets.
5. Various uncontrolled intellectual factors were also contributing to successful solution.
6. We used a highly select sample of subjects.
7. The three problems which we used, although different in nature, were hardly representative of typical problems.

Four general conclusions are suggested as a result of this experiment.

1. The extent of use of a particular method of attack varies with different problems and is dependent upon the kind of task involved.

2. On a particular problem, certain presumably different methods may be identical.

3. On a particular problem, certain so-called "methods" may be meaningless: mere terms having no meaningful reference to possible behavior.

4. The potential efficacy of a particular method of attack depends upon the kind of problem involved.

BIBLIOGRAPHY

1. BARLOW, M. C. Transfer of training in reasoning. *J. educ. Psychol.*, 1937, **28**, 122-128.
2. CHANT, S. N. F. Objective experiment on reasoning. *Amer. J. Psychol.*, 1933, **45**, 282-291.
3. CHEN, HAN-PIAO. Studies on Thurstone's Reasoning Tests. *J. genet. Psychol.*, 1937, **50**, 15-25.
4. COMSTOCK, C. On the relevance of imagery to the process of thought. *Amer. J. Psychol.*, 1921, **32**, 196-230.
5. DREVER, J. I. The pre-insight period in learning. *Brit. J. Psychol.*, 1934, **25**, 197-203.
6. DURKIN, H. E. Trial-and-error, gradual analysis, and sudden reorganization: An experimental study of problem solving. *Arch. Psychol.*, 1937, **30**, No. 210.
7. EIDENS, H. Experimentelle Untersuchungen über den Denkverlauf bei unmittelbaren Folgerungen. *Arch. f. d. ges. Psychol.*, 1929, **71**, 1-66. Abstracted by W. Hartman, *Psychol. Abst.*, 1930, **4**, item 95.
8. HEIDBREDER, E. An experimental study of thinking. *Arch. Psychol.*, 1924, **11**, No. 73. Pp. 165.
9. KINNAMAN, A. J. Mental life of two Macacus Rhesus monkeys in captivity. *Amer. J. Psychol.*, 1902, **13**, 98-148, 173-218.
10. LINDLEY, E. H. A study of puzzles. *Amer. J. Psychol.*, 1897, **8**, 431-493.
11. LINDWORSKY, J. *Das Schlussfolgernde Denken.* Germany: 1916. Quoted by M. F. Dunn in "Psychology of reasoning," *Cath. Univ. Amer. Stud. Psychol. Psychiat.*, 1926, **1**, No. 1.
12. LYON, D. O. *Memory and the learning process.* Baltimore: Warwick and York, 1917. Pp. 179.
13. MAIER, N. R. F. Reasoning in humans: I. On direction. *J. comp. Psychol.*, 1930, **10**, 115-143.
14. PEAR, T. H. The relevance of visual imagery to the process of thinking: I. *Brit. J. Psychol.* (Gen. Section), 1927, **18**, 1-14.
15. PETERSON, J. C. The higher mental processes in learning. *Psychol. Monogr.*, 1920, **28**, No. 129. Pp. 121.
16. RUGER, H. A. The psychology of efficiency. *Arch. Psychol.*, 1910, **2**. Pp. 88.
17. SEASHORE, R. H. The pyramid puzzle. *Amer. J. Psychol.*, 1938, **51**, 549-557.
18. STORRING, G. Experimentelle Untersuchungen über einfache Schlussprozesse. *Arch. f. d. ges. Psychol.*, 1908, **11**, 1-127. Reviewed by R. S. Clark in "An experimental study of thinking," *Arch. Psychol.*, 1922, **7**, No. 48.
19. THORNDIKE, E. L. Animal intelligence. *Psychol. Rev. Monog. Suppl.*, 1898, **2**, No. 8. Pp. 109.
20. WILHELM, W. Über Mittel und Fehler beim Schlieszen. *Arch. f. d. ges. Psychol.*, 1934, **92**, 175-192. Abstracted in *Psychol. Abstr.*, 1935, **9**, item 1101.

VOLUME 64
NUMBER 8

WHOLE NO. 314
1950

Psychological Monographs:
General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

The Validity of a Multiple-Choice Projective Test in Psychopathological Screening

By
MARTIN SINGER
Veterans Administration Hospital
Northport, Long Island, N.Y.

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the University of Pennsylvania

Accepted for publication, November 7, 1949

Price $1.00

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

COPYRIGHT, 1950, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION

ACKNOWLEDGEMENTS

THE WRITER wishes to express his deep gratitude to Professor Malcolm G. Preston for his invaluable aid in the formulation and execution of this study. Acknowledgements must also be made to Professor Robert A. Brotemarkle, chairman, and to the other members of the department of psychology, University of Pennsylvania, and to Drs. Richard H. Paynter, Morton L. Wadsworth and Earl E. Swartzlander of the Veterans Administration for their interest and help.

The necessary subjects for this study could not have been obtained without the cooperation of very many individuals from various institutions. Space limitations permit only the following token acknowledgements:

Brooklyn State Hospital: Drs. Clarence Bellinger, Nathan Beckenstein, Louis Mantone, James Rappa, Paul Tarantalla, Sidney Tamarin, and Joseph Villara.

Jewish Hospital of Brooklyn (Rehabilitation Clinic): Dr. Joseph L. Abrahamson and Mrs. Eugenia Gorman.

Mt. Sinai Hospital (Veterans Clinic): Dr. Paul Goolker, Mrs. Roslyn Braun, and Miss Ruth Cuker.

Post Graduate Hospital (Veterans Clinic): Dr. William D. Sherwood and Miss Norrie O'Brien.

Veterans Administration: Dr. Helen M. Campbell, Dr. Irwin J. Cohen, Mr. John W. Green, Mr. Robert Lackey, Dr. Bernard Locke, Dr. Harry M. Harter, Dr. James Q. Holsopple, Mr. Thomas MacNamara, Dr. Charles M. Morris, Mr. Martin Munzer, and Dr. Morris J. Tissenbaum.

Finally, thanks are due to Miss Margaret J. Crookhorn and the personnel in the Mental Hygiene Service of the New York Regional Office, Veterans Administration, for their kind assistance.

MARTIN SINGER

CONTENTS

I. INTRODUCTION 1
II. EXPERIMENTAL DESIGN 2
III. DESCRIPTION OF THE POPULATION TESTED 7
IV. TREATMENT OF THE DATA 11
V. RESULTS 14
VI. DISCUSSION OF THE RESULTS 27
VII. CONCLUSIONS 33
VIII. APPENDIX 34
IX. BIBLIOGRAPHY 39

THE VALIDITY OF A MULTIPLE-CHOICE PROJECTIVE TEST IN PSYCHOPATHOLOGICAL SCREENING[1]

I. INTRODUCTION

THERE is marked need for a test which can be rapidly administered and scored, and which has sufficient validity for use as a screening device in the detection of psychopathology. The present study is an attempt to determine whether the projective test approach[2] may be utilized for this purpose.

In the past decade, more and more clinical psychologists have turned to the use of projective tests in their attempts to identify psychopathology. Most of the projective tests have been standardized under conditions of individual administration and free association responses. Such tests have not been suitable as screening instruments since a great deal of time is required by a highly skilled psychologist to administer and interpret the results for each subject. However, there have been some projective tests in which the subjects selected their responses from several multiple choices listed, permitting rapid and objective scoring by untrained personnel.[3] The most widely used test of this type has been the Harrower Multiple Choice Rorschach (19). Inkblots were also employed as stimuli by Terman (52) and by Harrower and Grinker (4, 17). Pictures were used as stimuli by Clark (5), Franck (11), and Horowicz (23), Murray and Stein (38), Szondi (7, 41) and Temple and Amen (8, 51). Single words served as stimuli for Crown (6), Malamud (31), Maller (34), Mittelman and Brodman (36) and Pressey (39, 40), while story plots were used by Kelly and Bishop (27), Taylor (50) and Wolfenstein (59).

A review of the literature indicates that most of the multiple choice projective tests achieved little discrimination. In those cases where a fair amount of validity was reported, faults in the experimental design, such as improper cross validation, use of extreme groups, or testing of too few subjects, made the validity coefficients suspect. On the other hand, none of the studies which found no discrimination was sufficiently comprehensive to warrant a negative conclusion; namely, that the limitations inherent in any attempt to restrict responses to those listed on a test blank preclude valid discrimination with a projective test. Therefore, the present experiment was undertaken in an effort to determine the potentialities of the multiple choice projective test approach in the detection of psychopathology.

A multiple choice projective test was constructed and administered under what were considered "optimum" conditions; i.e., under those conditions which seem most likely to result in the detection of psychopathology. If, with such a test, a fair degree of validity were obtained, then further experimentation with this approach would seem indicated. On the other hand, if very little or no validity were obtained, then it would appear more profitable to explore other approaches—unless, of course, it was established that optimum conditions did not exist so that factors which might lead to radically different results were not included in the present experimental design.

---

[1] Sponsored by the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the author are a result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration.

[2] Comprehensive discussions of the projective methods may be found in Frank (12), Sargent (46) and White (57). Each of a series of reviews by Symonds and Samuel (49), Symonds and Krugman (48), and Hertz, Ellis and Symonds (21) has summarized the literature published since the preceding review. Briefly, the projective test approach is defined as one in which the subject is required to structure an ambiguous stimulus. In doing so, he is presumed to reveal his own personality structure.

[3] There have also been some attempts to quantify tests calling for free association responses, by organizing the responses into various categories which are scored linearly; e.g., the Rorschach Inspection Technique as originated by Munroe (37) and modified by Ross and Ross (45). It is felt that the validity of such tests as screening devices should be investigated if a completely objective scoring system has been proven inadequate.

II. EXPERIMENTAL DESIGN

Two areas where optimum conditions must exist may be distinguished: first, in the selection of the particular psychopathology to be identified, and second, in the test methods used to identify this psychopathology.

A. Type of Psychopathology Selected for Identification

In selecting the psychopathology to be identified, it is recognized that the grosser such psychopathology is, the more optimum conditions will be for valid identification. However, the grossness of the psychopathology must be limited so that useful conclusions may be drawn from the results. For example, the successful identification of deteriorated psychotics by a multiple choice projective test would be of little value. A screening device is not needed for this purpose, and no inferences could be drawn as to the probabilities of finer and more useful psychopathological identification. Therefore, the grossness of the psychopathology selected for identification in the present investigation was limited to such distinctions as would be required of a screening device.

Since projective tests have been used most successfully in the detection of psychopathological diagnoses, the attempt was made in the present investigation to differentiate among three diagnostic categories: normals, anxiety reactions, and schizophrenics. These particular categories were selected because they provided adequate distinctions and because their frequency made available a sufficiently large number of subjects in each category.[1] A more detailed consideration of these diagnostic categories will be presented in section III.

B. The Development of a Multiple-Choice Projective Test

The review of previous research with multiple choice projective tests disclosed no test with sufficiently promising results to warrant further exploration of that particular instrument. Therefore, a new test was constructed, based on the results obtained with individual, group, and multiple choice projective tests, and upon recommendations of various authorities in the field of projective testing. In this test, certain conditions were fixed since time considerations limited the length of the test. However, it was possible to vary other factors within the test so that the effect of such variation could be estimated.

1. Factors Not Varied

a. *Stimuli:* The Rorschach inkblots were selected as the stimuli because the Rorschach appears to be the best projective test for the prediction of psychopathology.

According to Rapaport and Schafer, "The use of the Rorschach Test, as one in a battery of tests, in a clinical-psychiatric setting for purposes of diagnoses and appraisal of adjustment shows . . . . that it is beyond doubt the most potent single diagnostic instrument clinical psychology possesses" (42, p. 73). Brody, in reviewing the literature on mental testing from 1938 to 1942, declared: "Increased use of mental tests for the study of psychoses and allied disorders is conspicuous in recent psychiatric practice. This increase is most prominent first in the use of the Rorschach ink blots" (3, p. 127). Symonds and Krugman point out the superiority of the Rorschach indirectly with their statement: "Next to the Rorschach, the Thematic Apperception Test has claimed most attention. Indeed, in numerous testing programs the Thematic Apperception Test and the Rorschach are listed as the two preferred methods for the study of personality. Actually, much yet remains to be done in exploration of the use of the Thematic Apperception method before it can receive general clinical use" (48, p. 88). Michael and Buhler (35) found the Rorschach to be superior as a diagnostic instrument to a variety of other tests both projective and non-projective.

---

[1] It was recognized that specific situations in which a test would be required to discriminate only among subjects in these three diagnostic classifications would rarely arise. However, the purpose of the experiment was *not to develop a test for practical use, but to determine whether the multiple choice projective test approach is capable of valid psychopathological differentiation.* If *potential validity* is established, then efforts can be made to develop a test for practical purposes; e.g., to screen inductees in the armed forces, to select employees in industry, or to discover students with marked personality problems. In order to limit the psychopathological deviation of subjects in this study to that which would exist in populations such as those mentioned above, the most seriously disturbed anxiety reactions and schizophrenics were not tested. (See section III.)

While in confirmatory experiments (2, 15, 26, 33, 47, 58) poor results were obtained with the Harrower Multiple Choice Rorschach, which used the Rorschach inkblots as stimuli, such results might well have been caused by inadequacies of the specific test rather than by the nature of the stimuli. Hunt and Stevenson have remarked that ". . . . had Harrower-Erickson followed the usual painstaking steps of item analysis and test validation, relying more on empirical investigation and less on a priori judgment in selecting her alternative responses and constructing her scoring key, the results might have been different" (24, p. 110).

b. *Method of administration:* As the purpose of the present study was to construct and evaluate a test which could be rapidly administered, it was thought desirable to follow a group administration procedure. The Rorschach slides were used since various investigators (18, 20, 22, 29, 30) have found that responses made to these slides are comparable to those made to the individual cards. These slides were projected on a screen in a semi-darkened room. An alternative possibility, the use of booklets in which the Rorschach cards were printed, was eliminated because of the expense of printing such booklets, and because the time of stimulus presentation would have been difficult to control, particularly with the non-normal subjects.

c. *The use of multiple choice questions:* It is only with some form of a multiple choice question that the requirement of rapid objective scoring can be met.

d. *The empirical determination of the scoring weights to be given each response:*[2] Results obtained with the Harrower Multiple Choice Rorschach by Due, Wright and Wright (9), Malamud and Malamud (31, 32), Lawshe and Forster (28), and Blair and Hoehn (2) have clearly indicated that a reponse to a multiple choice Rorschach might not have the same psychopathological significance as the same response to the free association Rorschach. Therefore, the psychopathological significance of multiple choice Rorschach responses must be determined independently.

2. *Factors Varied within the Test*

The factors which were varied are summarized below. They will be discussed more fully in subsequent sections.

a. *Structure of the questions:*

(1). *A "relative" versus an "absolute" decision by the subject.* In some questions the subject was required to select one of four percepts (relative decision), while in other questions the subject was required to rate a single percept on a five point rating scale (absolute decision).

(2). *Differences in the types of alternatives*

---

[2] At this point, it would seem advisable to define several terms which ordinarily may have different connotations.

A *percept* is a possible description of the whole slide or some part of the slide. It may be specific, such as "A bat"; abstract, such as "Death"; or a denial, such as "Nothing at all."

An *alternative* is one of the multiple choices listed in a question. It may be a percept, as when the subject is required to select one of four percepts listed, or it may be a rating, as when the subject is required to select one of five ratings on a rating scale.

A *response* is the answer a subject makes to a question. A subject's response can be either one of the alternatives listed in a question, or it may be a failure to select any alternative.

included in a question. The types of alternatives were varied to form the following question "designs."

(a). Design I (forced "normal" choice) contained what were considered on a priori grounds to be two percepts characteristic of normals, one percept characteristic of anxiety reactions, and one percept characteristic of schizophrenics.[3]

(b). Design II (unlimited choice) contained one "normal" percept, one "anxiety reaction" percept, one "schizophrenic" percept, and one percept worded "Something else."

(c). Design III (forced "abnormal" choice) contained two "anxiety reaction" percepts and two "schizophrenic" percepts.

b. *Time of stimulus presentation:* As subjects' immediate perceptions might differ from their more reflective perceptions, the time of stimulus presentation included both a brief exposure of the slides and an exposure which gave the subjects as much time as they required.

c. *Length of test:* The total test battery was made long enough so that information might be obtained as to the effect of increasing the size of the test by comparing the results obtained from sections of the test with those obtained from the total test.

d. *Method of determining scoring weights for each response:* While all methods depended upon the relative percentage of responses made by two groups, various formulae were employed which required varying amounts of difference between the two percentages for responses to be given scoring weights.

3. The Final Test Battery

Having determined the psychopathology to be identified (normal, anxiety reaction, and schizophrenia as discussed in section II, A) and the general composition of the test to be used for this identification, the next step was to construct a specific test. After experimentation with preliminary groups, the following test battery, termed the Singer Multiple Choice Rorschach (SMCR) to differentiate it from the Harrower Multiple Choice Rorschach, was developed. (A copy of the SMCR is shown in Appendix A.) The battery consisted of three tests.

a. *Test A:*

(1). *Number of questions:* 30

(2). *Number of percepts listed in a question:* 4

(3). *Type of percepts in a question:* There were 10 questions each of Designs I, II and III. (The three questions for each slide included one of each design. The order in which the designs were presented was varied systematically.)

(4). *Response by the subject:* The subject was required to select one of the four percepts in each question.

(5). *Time of stimulus presentation:* A slide was shown for 10 seconds for each question. Subjects were then allowed a few seconds more in which to record their response if they had not already done so.

b. *Test B:* This test was exactly the same as Test A with the following exceptions:

(1). *Different percepts were used:* However, the diagnostic significance of each percept paralleled that in Test A.

(2). *Time of stimulus presentations:* The subjects were allowed one and one-half minutes to answer the three questions for each slide. Preliminary experimentation had shown that when given an unlimited amount of time, over 95% of the subjects finished within the above time limit.

c. *Test C:*

(1). *Number of questions:* 100

(2). *Number of percepts listed in a question:* 1

(3). *Types of percepts in a question:* There were 10 questions for each slide, divided as follows:

---

[3] The diagnostic significance of the various percepts had to be determined on an a priori basis because no empirically established norms were available. Furthermore, it was necessary to base these a priori judgments on the results obtained with the free association Rorschach, since so little was known about multiple choice Rorschach responses.

The writer would like to express his appreciation to Drs. Bruno Klopfer and Zygmunt A. Piotrowski for their assistance in the selection of the percepts believed to be characteristic of the different diagnostic categories. The method by which the different percepts were chosen as characteristic of normals, anxiety reactions, and schizophrenics for the different inkblots is illustrated below by the way in which the "normal" percepts were selected for the first Rorschach blot.

The writer, after consultation with various experienced Rorschach examiners, drew up a list of about 15 percepts which were considered to be the most characteristic of normals. Drs. Klopfer and Piotrowski examined this list independently and each checked off the eight percepts they considered most characteristic. In some cases, they substituted new percepts as better choices than those listed. The writer used the selection of Drs. Klopfer and Piotrowski as a basis for choosing nine percepts for each card. Then, when the structure of a question called for a "normal" percept, one of these nine percepts was used. Which particular one was selected was determined in a relatively random fashion.

(a). Four "normal" percepts.

(b). Three "anxiety reaction" percepts.

(c). Three "schizophrenic" percepts, one of which was "Nothing at all."

(4). *Response by the subject:* Each question (percept) was rated by the subject on a five point absolute rating scale as follows:

(a). Exactly like it;

(b). Very much like it;

(c). Somewhat like it;

(d). Only a little like it;

(e). Not at all like it.

The term "like it" referred to how much either the whole slide or any part of it resembled the percepts listed.

(5). *Time of stimulus presentation:*

(a). Slide I. Three minutes.

(b). Slide II. Two and one-half minutes.

(c). Slides III through X. Two minutes.

The decreasing time limits were established because it was found that the subjects required less time to answer a question as they became familiar with the rating technique. Preliminary experimentation indicated that over 95% of the subjects finished within the above time limits when they were allowed unlimited time.

For Slide I subjects were warned when they had one minute left, while for the other slides they were informed when there were thirty seconds remaining.

The differences among the various tests may be observed more clearly in Table 1. The effect of these differences will be discussed in section IV.

4. Administration of the Test

The regular 3¼″ by 4¼″ Rorschach slides, purchased from the Psychological Corporation, were projected on a screen by a lantern projector illuminated by a lamp of at least 500 watts.

The screen image was generally at least 4′ by 5′ in size. In a few cases the image was slightly smaller because the size of the testing room made it impossible to focus a larger clear image. However, at such times the subjects were seated slightly nearer to the screen than in the cases where the 4′ by 5′ image was employed. It was possible, by having the light enter from the back of the room, to keep the image on the screen quite clear, while enough light existed for the subjects to read the alternatives on their test blanks. The room was kept in this semi-dark condition throughout the test. Subjects were seated at least 12 feet from the screen and at an angle of less than 30° from the center of the screen.

Before each test began, the subjects were given the answer sheets which listed the various possible responses for that test. They were told that slides would be

TABLE 1

CHARACTERISTICS OF THE DIFFERENT TESTS IN THE SMCR

Test	Time of Stimulus Exposure	Response Required of Subjects	Differences in Types of Percepts*						Total
			Design I		Design II		Design III		
			No. of Quest.	Percepts in a Quest.	No. of Quest.	Percepts in a Quest.	No. of Quest.	Percepts in a Quest.	
A	Brief	Select 1 of 4 percepts	10	N-2 AR-1 S-1	10	N-1 AR-1 S-1 "?"-1	10	AR-2 S-2	30
B	Relatively Unlimited	Select 1 of 4 percepts	10	N-2 AR-1 S-1	10	N-1 AR-1 S-1 "?"-1	10	AR-2 S-2	30
C	Relatively Unlimited	Rate each percept on an absolute rating scale.	40	N-1	30	AR-1	30	S-1	100

* N stands for a percept considered on an a priori basis to be characteristic of normals, AR for one characteristic of anxiety reactions, and S for one characteristic of schizophrenics. The "?" stands for the alternative "Something else."

projected on the screen, and that their task was to select a response while the slide was on the screen. No references were made to the Rorschach or to ink-blots. The specific procedure for each test was as follows:

a. *Test A*. The first Rorschach slide was projected on the screen for ten seconds, during which time the subjects checked one of the four alternatives in Question 1. The slide was taken off the screen and after a few seconds the same slide was again projected for 10 seconds while the subjects answered Question 2. After an interval of a few seconds, the slide was shown for another 10 seconds while the subject answered Question 3. The second Rorschach slide was then shown for 10 seconds with the subjects responding to Question 4. This procedure continued, with a new slide being used for each three questions, until all 30 questions were answered.

b. *Test B*. The first Rorschach slide was again projected on the screen while the subjects answered Questions 1A, 2A and 3A. The slide remained on the screen for 90 seconds while all three questions were answered. Then the second Rorschach slide was shown for 90 seconds while the subjects responded to the next three questions, and so on until all 30 questions were answered.

c. *Test C*. The first Rorschach slide was shown on the screen for three minutes while the subjects rated all 10 percepts for that slide. The second slide was projected for two and a half minutes while the subjects rated the second 10 percepts. Then the other slides were shown in order for two minutes each while the subjects rated the 10 percepts listed for each slide.

III. DESCRIPTION OF THE POPULATION TESTED

In the original experimental design, it was planned to establish the scoring weight for each response empirically from the results of 150 subjects, consisting of 50 normals, 50 anxiety reactions, and 50 schizophrenics, and to cross validate those weights on a second set of subjects. (The term "set" will always refer to a group consisting of 50 subjects in each of the three diagnostic categories.) However, since it proved impossible to obtain subjects in the various diagnostic categories who were thoroughly evaluated, and since the scoring weights established from only 50 subjects in each category proved highly unstable, a third set of subjects was tested to provide a larger sample upon which to establish scoring weights. After the testing had been completed, it was found possible to add a fourth set consisting of an "excess" of properly qualified subjects who were not needed for the third set, and subjects who had been excluded from previous sets because they did not quite meet all the requirements.

Over 2,000 subjects were tested, as it frequently was necessary, for administrative reasons, to include extraneous subjects in a testing group. The total number of groups tested was 90, of which 39 included only normals, 27 included only hospitalized schizophrenics, and the remaining 24 contained both anxiety reactions and out-patient schizophrenics. The results of 653 subjects, 250 normals, 203 anxiety reactions, and 200 schizophrenics, were used.

A. Age and Education

Information was obtained as to the age and education of the subjects, and the three diagnostic groups proved to be fairly homogeneous in these respects. The mean age of the normals was 25.0 years, that of the anxiety reactions was 30.0 years, and of the schizophrenics 28.4 years. Standard Deviations were 5.0, 6.6 and 6.4 respectively. There was a great deal of overlap among the groups with 93% of the normals, 87% of the anxiety reactions, and 87% of the schizophrenics falling between the ages of 20 and 35.

The educational level of the subjects was determined by the highest grade completed in school. Information was lacking for 5 normals, 6 anxiety reactions, and 21 schizophrenics, but there was no reason to believe that the educational level of these subjects was much different from that of the others. The mean highest grades reached were 13.2 for the normals, 11.4 for the anxiety reactions and 11.5 for the schizophrenics, with respective Standard Deviations of 2.0, 2.5 and 2.6. Seventy-eight per cent of the normals, 45% of the anxiety reactions, and 46% of the schizophrenics were graduated from high school. The percentages for those completing the second year of high school were 96%, 71% and 72% respectively. The few subjects who had less than an eighth grade education (6% of the anxiety reactions and 7% of the schizophrenics) were generally older or foreign born individuals.

No attempt was made to equate educational level, since the lower scholastic accomplishment is characteristic of the general anxiety reaction and schizophrenic population. It was felt that there was sufficient overlap among the three diagnostic groups to preclude any substantial differences arising merely as a reflection of variations in education. The possible effect of differences in age and amount of schooling has been evaluated in section V,B,8.

B. Description of Subjects in Each Diagnostic Category

The general characteristics of the subjects in each diagnostic category will be discussed first; then the composition of the four different sets will be presented.

1. *Normals*

Subjects in this group were drawn from veterans who came to the New York Regional Office of the Veterans Administration for a wide variety of nonpsychiatric reasons (e.g., to obtain information relating to insurance, benefits under the G.I. Bill of Rights, medical or dental treatment, etc.). All such veterans were first required to visit the Contact Division where they were interviewed and then routed to the appropriate section of the VA. Since the flow of veterans varied greatly during the week, at peak periods it was necessary for them to wait an hour or more before being interviewed. At this time, an announcement was made over a loud speaker asking male veterans of World War II to volunteer to take a personality test while they were waiting. All such veterans who took the test were identified merely by a number which they placed on each page of their test blank. This was done to insure honesty of response as much as possible.

Subjects were selected as normals on the basis of their responses to two questionnaires, one of the Cornell Selectee Index, Form N (53, 54, 55, 56), and the other a six-item questionnaire (listed in Appendix B) which was validated[1] against the Cornell Index, in an effort to include the best adjusted individuals in this group.

2. *Anxiety Reactions*

Subjects in this category were obtained from the VA Mental Hygiene Services of New York, Brooklyn and Philadelphia, and from the Rehabilitation Clinic of the Jewish Hospital of Brooklyn, the Mt. Sinai Hospital Veterans Clinic, and the Post Graduate Hospital Veterans Clinic. All subjects were out-patients and veterans of World War II.

Originally, only patients diagnosed by a psychiatrist who had treated them for at least several months were accepted in this group. However, it became impossible to obtain a sufficient number of subjects fulfilling this criterion. As the research progressed, there was a great decrease in the number of veterans diagnosed as anxiety reactions since, according to several psychiatrists consulted, those patients whose maladjustment arose primarily out of situational difficulties in the armed service or in becoming readjusted to civilian life either recovered or "bound" their anxiety in a specific neurotic or psychotic reaction pattern. Many patients who retained their free floating anxiety would be diagnosed by a psychiatrist, after a few months of treatment, in terms of the dynamic factors underlying the anxiety rather than as anxiety reactions.

Finally, for those patients still diagnosed as anxiety reactions, interrogation of the psychiatrists making these diagnoses generally revealed that this category was being used as a "waste basket" and that many of the patients had either a mixed psychoneurosis, or were possible latent schizophrenics.

Therefore, in order to secure the desired number of subjects, it became necessary to accept a much less rigid criterion; namely, the diagnosis of a screening psychiatrist who had interviewed the patient for about 45 minutes. It is recognized that many of the patients diagnosed by a screening psychiatrist as anxiety reactions might, after more intensive diagnostic work, be classified differently. However, none of these subjects may be considered normal in the sense of being well adjusted, relatively few could have been schizophrenic, and all had at least a fair amount of free floating anxiety.

---

[1] Only 10 of 233 subjects who answered all "No's" to this questionnaire, which was the normality criterion, scored over 14 on the Cornell Selectee Index, Form N, and only 34 of the 233 scored over 10.

3. *Schizophrenics*

About half the subjects in this group came from the Veterans Hospital at Northport, while most of the others, some of whom were also veterans of World War II, came from Brooklyn State Hospital. A small number were being treated as out-patients in the New York and Brooklyn Mental Hygiene Services of the VA.

For the most part, the subjects were in relatively good condition. Not only were the best patients selected from the two hospitals, but the patients had to be willing and able to fill out a fairly complicated test satisfactorily. ("Satisfactory" standards are described in the footnote in Section IV,B,2.) No coercion was used to make the subjects take the test, and even less aid was given them than under many group testing situations, since no proctor assistance was made available during the test.

Although question concerning the original diagnosis of schizophrenia existed in less than a half dozen subjects, many of the patients were considered at the time of testing to be at or near remission. Almost half the Northport subjects were open ward patients, and many were released or sent out on trial visits less than two weeks after taking the test. Thus, subjects in this category represented the "best" schizophrenic population; that is, the type of subject who often might not be identified in a brief interview or from responses to a questionnaire.[2]

No attempt was made to use the schizophrenic sub-classifications (catatonic, hebephrenic, simple, and paranoid) because nearly all psychiatrists felt that the sub-classifications they gave to many of the subjects were unreliable.

4. *Special Characteristics of the Four Sets of Subjects*

a. *Set C1.* While each set of subjects was supposed to include 50 persons in each of the three diagnostic categories, there were actually 100 normals and 53 anxiety reactions and 50 schizophrenics in this set because some of the subjects did not complete the entire battery. However, the combined results of the subjects in each category was equivalent to results which would have been obtained had there been 50 subjects in each category completing the test battery.

(1). *Normals.* These subjects were available for but half an hour. Therefore, they were screened with only the six-item questionnaire, and 50 were given Tests A and B while another 50 were given Test C.

(2). *Anxiety Reactions.* Most of these subjects were thoroughly evaluated by a psychiatrist. There were 53 subjects, 41 of whom took all three tests. One subject took only Test A, two took Tests A and B, two took Tests A and C, and seven took Tests B and C. These discrepancies occurred because some subjects did not answer all the tests satisfactorily, while others came in late and missed part or all of Test A. Results were available for 50 subjects on Tests B and C, but for only 46 subjects on Test A. Therefore, the frequencies in the intervals in the distributions of the last group were multiplied by 50/46 to obtain comparable distributions.

(3). *Schizophrenics.* All 50 subjects took the complete test battery. Forty-two of them were patients at the Northport Veterans Hospital while the other eight were out-patients.

b. *Set C2.* There were 50 subjects in each diagnostic category in this set and in each of the two following sets. All subjects completed the entire test battery.

---

[2] A survey of the 110 Northport subjects, made from one to two years after testing, revealed that 33 had been discharged, 17 were out on trial visit status, 2 were AWOL (absent without leave), 5 had been transferred to other hospitals, and 53 remained. Thus, about half no longer required hospitalization. Of those remaining, many have been in good enough condition to be sent out on trial visit or leave of absence since their testing.

(1). *Normals.* These subjects were first screened by the six-item questionnaire, and were given the Cornell Selectee Index, Form N, after completing the SMCR. Only subjects with a score of 10 or below on the Cornell Index were included in this set.

(2). *Anxiety Reactions.* There were 29 subjects selected on the basis of a screening diagnosis. The other 21 subjects were evaluated by the psychiatrist treating them. However, in many cases, the psychiatrist had not seen the patient long enough to formulate more than a tentative diagnostic impression.

(3). *Schizophrenics.* There were 25 subjects from Northport Hospital, and 25 from Brooklyn State Hospital.

c. *Set X.*

(1). *Normals.* These subjects were selected as in Set C2.

(2). *Anxiety Reactions.* There were 41 subjects selected on the basis of a screening diagnosis, while the other 9 subjects were given an evaluation, in some cases a tentative one, by the psychiatrist treating them.

(3). *Schizophrenics.* This group included 25 subjects from Northport, and 25 from Brooklyn State Hospital.

d. *Set C3.* This set was composed of subjects who were in excess after the quota for Set X had been filled, plus subjects who had not quite met the requirements for the previous sets.

(1). *Normals.* There were 10 subjects who met the requirements outlined in C2; 9 subjects who scored 10 or below on the Cornell Index but who had answered one "yes" on the six-item questionnaire; 28 subjects who had scored between 11 and 14 inclusive on the Cornell, 8 of whom answered one "yes" and one of whom answered two "yes's" to the six-item questionnaire; and three subjects who completed only half of the Cornell Index but whose score on this half indicated that they would probably have scored under 10 had they completed the test.

(2). *Anxiety Reactions.* There were 33 "excess" subjects, and 17 subjects who had a supplementary diagnosis, such as "and depression" or "in an inadequate personality" which indicated that they were not relatively pure anxiety reactions. There were 18 in the former group and 11 in the latter whose diagnosis was made by a screening psychiatrist.

(3). *Schizophrenics.* This group included 9 out-patients, 18 open ward patients from Northport Hospital and 23 patients in the process of insulin shock treatment at Brooklyn State Hospital.

Sets C1, C2 and C3 were used as criterion populations upon whom scoring weights were established, while Set X served as the cross validation population.

IV. TREATMENT OF THE DATA

A. DIAGNOSTIC GROUPS COMPARED

IN ORDER to demonstrate that psychopathology can be detected by a multiple choice projective test, it would be necessary to discriminate successfully only between two of the three diagnostic groups tested. Therefore, in the treatment of the data, each pair of two diagnostic categories was considered separately. Thus, a comparison was made between the normals and the anxiety reactions; another comparison was made between the normals and the schizophrenics; and a third comparison was made between the anxiety reactions and the schizophrenics. For convenient reference, each set of two diagnostic groups was termed a "comparison pair," resulting in:

1. A normal-anxiety reaction comparison pair (N-AR)
2. A normal-schizophrenic comparison pair (N-S)
3. An anxiety reaction-schizophrenic comparison pair (AR-S)

Each of these comparison pairs was treated as a distinct unit, and the procedure outlined below for the N-AR pair was duplicated for the other two.

B. STATISTICAL PROCEDURE

1. *Method of Determining Scoring Weights*

The scoring weight given each response was determined by the relative percentage of subjects in each diagnostic group selecting that response. The procedure followed is outlined below.

a. The percentage of normals and anxiety reactions making a response was computed. Possible responses included the various alternatives listed in a question plus the failure to select any alternative.

b. The significance of the difference between these percentages was then determined according to the formulae outlined in paragraph c. below.

(1). If the normals' percentage was significantly higher, the response was given a weight of minus 1.

(2). If the anxiety reactions' percentage was significantly higher, the response was given a weight of plus 1.

(3). If the difference was not significant, the response was weighted 0.

(In the N-S comparison pair, a significantly higher normal percentage received a minus 1 weight, while in the AR-S comparison pair, a significantly higher anxiety reaction percentage received a minus 1 weight.)

c. Two different formulae were used to determine the significance of the difference between the percentages.

(1). The first was a formula adapted from one outlined by Guilford[1] (14, pp. 298-301):

$$W = \frac{P_1 - P_2}{pq},$$

where W = *Weight*. When the weight came to plus 0.5 or above, a weight of plus 1 was given. Similarly, a weight of minus 0.5 or below was weighted minus 1.

P_1 = the percentage of subjects in group 1 making a response.

P_2 = the percentage of subjects in group 2 making a response.

$p = \frac{P_1 + P_2}{2}$

$q = 1 - p$

(2). The second was the customary formula for finding the significance of a difference between percentages, namely:

$$C.R. = \frac{P_1 - P_2}{\sqrt{\frac{P_1Q_1}{N_1} + \frac{P_2Q_2}{N_2}}},$$

---

[1] In Guilford's formula, a constant, plus 4, is added so that possible weights will all be positive and range from 0 through 8. Guilford's formula was modified because the writer found it more convenient to work with plus and minus weights. Weights were restricted to plus or minus 1 as the very few responses which would have received higher weights with the Guilford formula would have caused no appreciable difference in the results.

where C.R. = the critical ratio, and the other symbols have their conventional meaning.

With this formula, the determination of whether a difference is significant will depend upon the size of the critical ratio required. The higher the critical ratio required, the greater the difference between the two percentages will have to be. Four different critical ratios were used as criteria to determine weights as follows:

- (a) A C.R. of 1.6449, which is equivalent to a p of .10.
- (b) A C.R. of 1.9600, equivalent to a p of .05.
- (c) A C.R. of 2.3263, equivalent to a p of .02.
- (d) A C.R. of 2.5758, equivalent to a p of .01.

d. Thus, there were five different weighting "systems" used:

- (1). That based on the formula of Guilford (referred to in future as weighting system G).
- (2). The four based on the regular percentage-difference formula where the p required to indicate a significant difference varied as follows:
  - (a) $p = .10$ (referred to as weighting system .10).
  - (b) $p = .05$ (referred to as weighting system .05).
  - (c) $p = .02$ (referred to as weighting system .02).
  - (d) $p = .01$ (referred to as weighting system .01).

2. *The Responses Weighted*

Since failure to respond to a stimulus is frequently a psychopathological sign indicative of "stimulus shock" or evasion on the individual Rorschach, such failures were treated as regular responses. Thus, in Tests A and B where the subjects were requested to select one of four percepts, there were five possible responses; four of them were the different percepts listed, while the fifth was the failure to select a percept. Similarly, in Test C, the five steps in the rating scale gave rise to five possible responses, while failure to select any rating provided a sixth possible response. The relatively few cases in which the responses were "spoiled" by the checking of more than one percept or rating were scored as "failure to make a selection."

To make sure that failure to make a selection did not represent a general negativism toward the test situation, the results of the subjects who had selected too few of the alternatives listed were eliminated.[2]

3. *Criterion Populations*

Scoring weights were determined from several criterion populations as follows:

a. *Set C1.* Weights were established upon this population for the purpose of scoring the same population. It was recognized that such treatment would lead to a spuriously high differentiation since all "chance" differences were weighted in favor of greater differentiation. However, if little differentiation resulted from this treatment, then cross validation would be unnecessary, as no better results could be expected.

b. *Sets C1 and C2 were combined to form population C12.* This was the criterion population upon which weights were developed for cross validation on Set X.

c. *Sets C1, C2 and C3 were combined forming population C123.* This was also used as a criterion population upon which weights were developed for cross validation on Set X. Population C123 was used to obtain information regarding the effect of:

(1). Increasing the size of the criterion population.

(2). Increasing the size of the critical ratio required for a response to be given a scoring weight. There were too few responses given weights in criterion population C12 when a p of .02 or .01 was used as a standard to permit an adequate judgment concerning these standards.

---

[2] Except for a few anxiety reactions in Set C1, subjects were required to select specific alternatives in at least 60% of the questions in each test, and in 75% of the test battery as a whole. In the total population of 653 subjects, only 6 subjects responded properly to less than 80% of the questions in the battery, only 13 responded properly to less than 85%, and only 29 responded properly to less than 90%. Furthermore, whenever it was felt that a subject had made his selections capriciously, as was suspected when the response checks formed an arithmetic or geometric pattern, no data from this subject were tabulated or scored.

4. *Weighting Systems Used with Different Criterion Populations*

a. With population C1, only weighting system G was used.

b. With population C12, the weighting systems used were:

(1). G
(2). $p = .10$
(3). $p = .05$

c. With population C123, the weighting systems used were:

(1). G
(2). $p = .10$
(3). $p = .05$
(4). $p = .02$
(5). $p = .01$

5. *Cross Validation*

All cross validation was done on population X. The statistics used to evaluate the differences found between two distributions were:

a. *The percentage of subjects identified incorrectly at the best critical score.* Normal subjects scoring above the critical score were considered incorrectly identified, as were anxiety reactions with scores below it. The critical score was placed at that point which would result in the minimum percentage of subjects incorrectly identified in the total population; that is, the minimum number obtained when the number of anxiety reactions incorrectly identified was added to the number of normals incorrectly identified.

Two "best" critical scores were found empirically, one based on the distributions of the criterion population, and the other based on the distributions of the cross validation population. The percentage of subjects incorrectly identified in the cross validation population at each of these critical scores was determined. It is recognized that the exact percentage of subjects incorrectly identified at a critical score will not be highly reliable. However, an indication of the possible variability can be obtained by noting the percentage of subjects incorrectly identified at a point a few scores higher and a few scores lower than the critical score.

Furthermore, the most desirable critical score would vary greatly with the purpose for which a screening measure was used. As Ellis and Conrad (10) have pointed out, in most practical situations the percentage of false negatives or the percentage of false positives, as the case may be, is of greater importance than a combined percentage of wrong identifications. Thus, a much higher critical score would be employed when the purpose of the test was to definitely eliminate anxiety reactions than when the test was used to select possible anxiety reactions for a more thorough evaluation. However, the present study is attempting to determine generally whether any psychopathological screening is possible with a multiple choice projective test, and consequently it is concerned with the largest amount of differentiation possible, rather than with the direction of the error.

b. *The significance of the difference between the two distributions.* This was determined as a critical ratio.

V. RESULTS

A. Preliminary Analysis upon Population C₁

After population C1 had been tested, a preliminary analysis of the data was made to determine whether further testing under the same experimental conditions was indicated. Scoring weights were developed from the responses made by the subjects in C1 using the Guilford formula (weighting system G). Then these weights were used to score the subjects in C1.

Since half of the "normal" subjects in C1 took only Test AB, while the other half took only Test C, it was not possible to obtain a full battery score for these subjects. Therefore, the results of Test AB and Test C were treated separately.

A comparison of the Test AB scores received by the normals and the anxiety reactions revealed that only 9% were incorrectly identified at the best critical score. For Test C, only 6% were incorrectly identified. In the normal-schizophrenic comparisons, only 8% were incorrectly identified with each of the two tests. An anxiety reaction-schizophrenic comparison was made only for Test AB where 12.5% of the subjects were incorrectly identified. In all comparisons, the critical ratios were over 10. These results appeared definitely to warrant further testing with the SMCR.

An analysis was made of the results obtained with Tests A, B and C separately, and with various sections within a test, to determine whether any modifications should be made in the battery before further testing was undertaken. However, the differences in test conditions which were investigated, such as in the time of stimulus exposure, appeared to produce little difference in the results obtained so that no changes were made in the test battery.

B. Cross Validation of the SMCR

In the data presented below, unless otherwise specified, the results will always pertain to the subjects in population X; that is, the cross validation population.

1. *Discrimination Obtained with Different Weighting Systems*

a. *Percentage of subjects incorrectly identified.* Table 2 presents the scores obtained by normal and anxiety reactions subjects when different sets of weights are used.

In this table, as elsewhere, N represents the normal subjects and AR the anxiety reactions. Under "Score" are presented the lower scores in each step interval. To the right of each score are the number of subjects found in that step interval. At the bottom of the page, "M" "SD," and "CR" have their conventional meanings. The "% Wrong" stands for the percentage of subjects incorrectly identified at the best critical score.

Tables 3 and 4, presenting the results of the normal-schizophrenic and anxiety reaction-schizophrenic comparison pairs, are similar in form to Table 2.

The amount of discrimination obtained for the three comparison pairs with different sets of weights is summarized in Tables 5A and 5B which present the percentage of subjects incorrectly identified (% Wrong) and the critical ratios respectively.

As can be most easily observed from the summary tables, 5A and 5B, the various sets of weights produce a remark-

TABLE 2

RESULTS OBTAINED WHEN THE NORMAL AND ANXIETY REACTION CROSS-VALIDATION POPULATIONS ARE SCORED WITH DIFFERENT SETS OF WEIGHTS

Criterion	Weights Based on C 12						Weights Based on C 123									
	Weighting System						Weighting System									
Score	G		.10		.05		G		.10		.05		.02		.01	
	N	AR	N	AR	N	AR	N	AR	N	AR	N	AR	N	AR	N	AR
20								2		4						
16	1						2	3		3	1	3				
12		3	1	2			1	1	3	2		4		1		
8		5	2	9	2	5	2	6	1	6	2	4	2	8	1	5
4	2	5		8	1	16	2	13	3	7	5	10	4	7	5	8
0	2	9	3	8	9	10	7	7	8	10	11	13	16	20	13	25
−1	7	6	14	11	21	13	13	8	10	5	11	7	22	9	30	10
−5	8	11	14	6	17	5	14	6	5	5	15	6	6	4	1	2
−9	14	4	11	5		1	7	2	11	5	5	2		1		
−13	8	4	4	1			2	2	9	2		1				
−17	4	2	1													
−21	3	1								1						
−25	1															
M*	−9.4	−2.3	−5.7	0.4	−2.4	1.3	−2.9	2.5	−4.8	2.4	−1.8	2.6	−0.5	2.0	−0.5	2.1
SD*	7.8	8.6	6.2	7.1	3.8	5.0	7.3	8.5	7.9	10.1	5.8	7.0	3.8	5.0	2.9	3.9
CR*	4.3		4.5		4.1		3.4		3.9		3.4		2.7		3.8	
% Wrong†	28		29		31		31		30		33		30		31	

---

* These calculations were made from data grouped with a class interval of 2. Because such a table was too large for one printed page, the class intervals were enlarged to 4 in the above table.

† These calculations were made from ungrouped data because in this statistic, unlike the others, grouping errors will occur in only one direction; that is, to increase the % Wrong.

TABLE 3

RESULTS OBTAINED WHEN THE NORMAL AND SCHIZOPHRENIC CROSS-VALIDATION POPULATIONS ARE SCORED WITH DIFFERENT SETS OF WEIGHTS

Criterion	Weights Based on C 12						Weights Based on C 123									
	Weighting System						Weighting System									
Score	G		.10		.05		G		.10		.05		.02		.01	
	N	S	N	S	N	S	N	S	N	S	N	S	N	S	N	S
42								1								
36		1						4								
30		2								1						
24		1		1						1		2				
18		4		2			1	1		2		2		1		
12	1	1		1		1	1	1		3				3	1	2
6		2	1	6		3	2	8	1	1	1	4	1	3	1	3
0	3	10	1	3	2	6		11		7	1	9	1	9		15
−1	3	10	2	10		10	5	8	2	9	4	10	4	12	7	10
−7	2	5	4	10	6	14	2	7	4	7	3	6	2	10	6	12
−13	5	7	6	8	4	7	3	4	4	5	3	9	11	6	20	6
−19	9	3	6	3	16	8	10	3	5	7	11	4	20	5	11	1
−25	9	2	10	5	18	1	13	1	7	5	16	3	9		4	1
−31	9	2	12	1	3		9	1	8		6	1	2	1		
−37	7		6		1		2		12	1	4					
−43	1		1				2		5	1	1					
−49	1		1						1							
−55									1							
M*	−23.7	−2.2	−25.0	−7.2	−21.9	−8.1	−21.6	0.9	−29.7	−7.9	−23.3	−5.8	−18.5	−5.2	−14.4	−3.9
SD*	13.4	16.1	12.1	13.4	7.6	9.5	14.3	16.5	14.0	16.3	11.6	13.2	8.9	10.5	8.3	8.7
CR*	7.1		6.9		7.9		7.2		7.1		7.0		6.8		6.1	
%Wrong†	20		22		18		18		21		20		19		22	

---

* These calculations were made from data grouped with a class interval of 3. Because such a table was too large for one printed page, the class intervals were enlarged to 6 in the above table.

† The calculations were made from ungrouped data because in this statistic, unlike the others, grouping errors will occur in only one direction, that is, to increase the % Wrong.

TABLE 4

RESULTS OBTAINED WHEN THE ANXIETY REACTION AND SCHIZOPHRENIC CROSS-VALIDATION POPULATIONS ARE SCORED WITH DIFFERENT SETS OF WEIGHTS

Criterion	Weights Based on C 12						Weights Based on C 123									
	Weighting System						Weighting System									
Score	G		.10		.05		G		.10		.05		.02		.01	
	AR	S	AR	S	AR	S	AR	S	AR	S	AR	S	AR	S	AR	S
20		1						1								
16		3						3		1						
12		3		3				1		1						
8	2	5	1	5		1	1	5				4		1		
4	1	4	1	5		5	3	5	3	7		3		3	1	5
0	4	3	3	4	2	8	1	2	1	3	3	6	6	14	4	17
−1	2	8	5	10	7	11	7	12	4	8	4	10	4	13	15	15
−5	2	9	2	9	9	12	3	7		7	6	10	14	11	17	10
−9	7	5	8	6	6	5	8	6	5	9	9	9	18	4	13	3
−13	6	3	12	2	17	5	6	2	8	7	11	2	8	4		
−17	6	1	8	2	9	2	11	2	7	1	13	5				
−21	10	3	6	3		1	8	3	9	1	4	1				
−25	3	1	4				2	1	7	4						
−29	5			1					5							
−33	2	1							1	1						
N*	−15.4	−2.7	−13.2	−3.8	−11.0	−4.6	−12.3	−2.6	−17.3	−7.4	−12.7	−5.3	−8.1	−2.9	−5.7	−1.7
SD*	11.2	12.6	8.9	10.7	6.0	8.3	8.9	11.1	10.1	10.4	6.4	7.8	4.6	5.5	3.9	3.9
CR*	5.3		4.7		4.3		4.8		4.8		5.2		5.1		5.1	
% Wrong†	24		24		27		27		24		26		25		31	

* These calculations were made from data grouped with a class interval of 2. Because such a table was too large for one printed page, the class intervals were enlarged to 4 in the above table.

† These calculations were made from ungrouped data because in this statistic, unlike the others, grouping errors will occur in only one direction, that is, to increase the % Wrong.

TABLE 5A

SUMMARY OF PERCENTAGE OF SUBJECTS INCORRECTLY IDENTIFIED (% WRONG) IN THE CROSS-VALIDATION SAMPLE

Comp. Pair	Weights Based on Population C 12			Weights Based on Population C 123					Mean C 12	Mean C 123
	G	.10	.05	G	.10	.05	.02	.01		
N-AR	28	29	31	31	30	33	30	31	29	31
N-S	20	22	18	18	21	20	19	22	20	20
AR-S	24	24	27	27	24	26	25	31	25	27
Mean	24	25	25	25	25	26	25	28	25	26

TABLE 5B

CRITICAL RATIOS FOR STATISTICS IN TABLE 5A

Comp. Pair	Population C 12			Population C 123					Mean C 12	Mean C 123
	G	.10	.05	G	.10	.05	.02	.01		
N-AR	4.3	4.5	4.1	3.4	3.9	3.4	2.7	3.8	4.3	3.4
N-S	7.1	6.9	7.9	7.2	7.1	7.0	6.8	6.1	7.3	6.8
AR-S	5.3	4.7	4.3	4.8	4.8	5.2	5.1	5.1	4.8	5.0
Mean	5.6	5.4	5.4	5.1	5.3	5.2	4.9	5.0	5.5	5.1

ably similar amount of discrimination. For each comparison pair, the middle six (of eight) "% Wrong" figures have a range of only three percentage points, while the critical ratio range is less than one point. This consistency clearly indicates that differences both in the weighting systems and in the criterion populations upon which the weights are based have had little effect upon the amount of discrimination produced.

b. *Particular subjects incorrectly identified.* It is possible that while the same percentage of subjects was incorrectly identified by each set of weights, there may be differences with respect to the particular subjects who were incorrectly identified. Therefore, a comparison of the subjects incorrectly identified by each set of weights was made.

For each diagnostic group in a comparison pair, there were eight distributions, a different one produced by each set of weights. The 10 subjects scoring lowest or highest, whichever indicated incorrect identification, were selected from each distribution as the ones most clearly misidentified. Then, for each diagnostic group in a comparison pair, the 10 subjects appearing in most distributions as incorrectly identified were considered the criterion "incorrectly identified groups." It was found that there was very little difference between the 10 subjects selected by any of the sets of weights and the subjects in the criterion group.

Because of the marked similarity in the results obtained with the various sets of weights, it was felt justifiable to use only one set in the subsequent treatment of the data. The weighting system $p = .10$ with weights derived from population C123 was selected because it identified so many of the subjects in the incorrectly identified group and because it led to the widest range in scores, an important factor when the results obtained from only sections of the test battery were considered.

2. *Effect of Increasing the Number of Subjects*

The fact that adding the subjects in population C3 to the criterion population did not increase the validity may be attributed to one of two reasons. Either the 100 subjects in population C12 provided a sufficiently large number to attain the best possible set of weights, or the poorer diagnostic classification of the subjects in C3 was enough to balance any additional validity to be gained from adding 50 subjects to each category in the criterion group. To investigate this problem, populations C2 and C3 were scored separately with weighting system $p = .10$ based on criterion population C123. Population C1 was not scored because a total score could not be obtained for the "normal" subjects since each of them had taken only half of the testing battery.

In the N-AR comparison pair, the percentage of subjects incorrectly identified was 18% for C2 and 16% for C3. In the N-S comparison pair, these percentages were 16% and 17% respectively, while for the AR-S comparison pair they were 19% and 25% respectively. While it is true that C2 and C3 each formed one-third of the criterion population upon which the weights were based, still if the diagnostic classification of the subjects in C3 was materially inferior, this population should have shown much less discrimination. Since the percentage of subjects incorrectly identified was very similar for both C2 and C3, it would seem more logical to conclude that for this test the diagnostic classification of subjects in C3 could not be much inferior to that of the subjects in C2. Hence, for the SMCR, the addition of subjects to the criterion population beyond 100 per diagnostic category would not seem likely to materially increase the amount of discrimination produced.

3. *Reliability of the SMCR*

Because subjects were not available for retesting, the reliability was determined by a measure of internal consistency. The test battery was divided into four sections: Test A, Test B, the first half of Test C, and the second half of Test C. Each section was then split into two halves according to a table of random numbers. Scores on one half of each section were summed to give a Part 1 score which was correlated against a Part 2 score derived from the sum of scores on the other half of each section.

For each of the three comparison pairs, reliabilities were found for each diagnostic group separately, and then for both groups combined. In the N-AR comparison, the reliability coefficients, when corrected by the Spearman-Brown prophecy formula, were .69 ± .07 for the normals alone, .75 ± .06 for the anxiety reactions alone, and .77 ± .04 when both groups were combined. In the N-S comparison, the corrected coefficients were .78 ± .05 (N), .82 ± .05 (S), and .87 ± .02 (N-S); while those for the AR-S comparison were .65 ± .08 (AR), .67 ± .08 (S), and .72±.05 (AR-S). These results would indicate that the SMCR has a reasonable degree of reliability.

4. *Validity of the SMCR*

There would seem to be no question that valid group differentiations have been made for each comparison pair. In Table 5B only one of the critical ratios was under 3.0, and that came to 2.7. However, the amount of error when individual prediction is desired can be gathered only from the percentage of subjects incorrectly identified. In Table 5A, when weighting systems $p=.01$ was eliminated because it was the only one which gave somewhat atypical results, the median of the other seven weighting sets indicate that 30% of the subjects in the N-AR comparison pair are incorrectly identified, as are 20% in the N-S comparison pair and 25% in the AR-S comparison pair.

However, it must be recognized that these percentages are based on the "best" critical score of the cross validation distributions. This procedure weights all sampling errors in favor of a greater discrimination, and thus gives a spuriously high validity. To properly estimate validity, the critical score should be based on a previous population. Therefore, the best critical score was found from the distributions of population C23 scored with weighting system $p=.10$ based on criterion population C123. These critical scores proved to be +2 for the N-AR comparison pair, −21 for the N-S comparison pair, and either −7 or −8 for the AR-S comparison pair. There were two possible critical scores for the last pair because both gave the same percentage of subjects incorrectly identified.

At the above cutting scores there would be 30% of the N-AR and 25% of the N-S comparison pairs incorrectly identified. Since a cutting score of −8 would result in 31% of the subjects being incorrectly identified, while a score of −7 would result in 29% of the subjects being incorrectly identified, it would appear best to estimate the percentage of subjects incorrectly identified in the AR-S comparison pair at 30%.

It should be recognized that using the results of only one rather than a series of samples to establish the critical scores, and having too few subjects in the distributions for smooth curves may lead to an underestimate of the validity. Therefore, the best possible estimate of the validity may lie somewhere between the above results and those obtained when using the best critical score of the cross validation distribution. However, exactly where in between these points it would lie is impossible to determine, and it would appear safest to conclude that the SMCR will incorrectly identify at most 30% of the N-AR comparison pair, 25% of the N-S comparison pair, and 30% of the AR-S comparison pair.

5. *Classifying Subjects According to Tri-Scores*

In the above estimate of validity, the subjects in a comparison pair were scored only with the set of weights derived from the distributions of that comparison pair. However, these subjects could also be scored with the weights derived from the other two comparison pairs. Thus, it was possible to score each subject with three sets of weights, one from each comparison pair, and thereby classify him:

1. As normal (N) or as anxiety reaction (AR) from the results obtained with the N-AR comparison pair scoring key.

2. As N or S from the N−S comparison pair scoring key.

3. As AR or S from the AR-S comparison pair scoring key.

Each subject could then be classified in one of the following eight ways: (These classification combinations will be referred to in the future as tri-scores.)

1. As N on the N-AR key; N on the N-S key; and AR on the AR-S key. NNA
2. N (N-AR key); N (N-S key); S (AR-S key). NNS
3. N (N-AR key); S (N-S key); AR (AR-S key). NSA
4. N (N-AR key); S (N-S key); S (AR-S key). NSS
5. AR (N-AR key); N (N-S key); AR (AR-S key). ANA
6. AR (N-AR key); N (N-S key); S (AR-S key). ANS
7. AR (N-AR key); S (N-S key); AR (AR-S key). ASA
8. AR (N-AR key); S (N-S key); S (AR-S key). ASS

To determine the validity obtained by using these tri-scores, the same validation procedure was followed as when subjects in a comparison pair were scored only with the set of weights derived from that comparison pair.

Population C2 was used as the criterion population, and the tri-scores of this group were found. The diagnostic identification of these subjects was based on whether they scored above or below a critical score. For each comparison pair the critical score was determined from the distributions of the two diagnostic groups within that comparison pair. (In all comparison pairs, the point at which the critical score fell proved to be the same for population C2 as had been found for population C23.)

For the cross validation population, two separate tri-score distributions were obtained; one where the diagnostic classification of the subjects was based on a critical score derived from the distribution of the cross validation population itself (Pop. X), and the other where the diagnostic classification was based on the point selected as the critical score in population C2. (Pop. X′). The tri-score distributions are presented in Table 6.

The tri-score weights were determined for each comparison pair by the tri-score

TABLE 6

NUMBER OF SUBJECTS IN EACH TRI-SCORE CLASSIFICATION

Tri-Score	Population C2 (Classified according to own "best" critical score)			Population X (Classified according to own "best" critical score)			Population X′ (Classified according to C2 critical score)		
	N	AR	S	N	AR	S	N	AR	S
NNA	39	12	5	29	14	3	32	19	8
NNS	0	1	0	1	1	0	1	0	1
NSA	1	1	2	4	5	4	6	1	6
NSS	6	0	14	6	0	20	1	0	12
ANA	1	8	0	2	6	0	4	9	3
ANS	0	0	1	0	1	0	0	1	0
ASA	0	20	3	3	11	3	3	12	7
ASS	3	8	25	5	12	20	3	8	13

distributions of the two diagnostic groups in a comparison pair. For example, in the N-AR comparison pair, the tri-score NNA was considered as a normal score since the population C2 section of Table 8 reveals that 39 normals and only 12 anxiety reactions made this tri-score. For two tri-scores, NSA and ANS, there were the same number of normals and anxiety reactions making these scores. In such cases, the tri-score was weighted according to the results of the comparison pair involving the two groups concerned. Thus, in the N-AR comparison, the results obtained with the N-AR set of weights would determine whether a subject would be classified as normal or as anxiety reaction.

It can be seen in Table 7 that for both group X and X′, the % Wrong using only the critical score in a subject's comparison pair gives about the same results as the % Wrong using the tri-score. The very limited improvement which results when the best possible combination of tri-scores is found for each group indicates that little could be expected from using a criterion population larger than C2 in establishing the tri-score weights.

However, the tri-score is not entirely without advantage. Thus, for subjects receiving the tri-score NNA, a statement could be made that they are most likely to be normals; but that if they are not normals, then they are more likely to be anxiety reactions rather than schizophrenics. There are a sufficient number of subjects with tri-scores NNS, ANA, ASA, and ASS to make similar interpretations possible for subjects in these clas-

TABLE 7

PERCENTAGE OF WRONG DISCRIMINATIONS OBTAINED WITH CRITICAL SCORES ONLY, AND WITH TRI-SCORES

	Population C2			Population X			Population X′		
	N	AR	S	N	AR	S	N	AR	S
With critical score only	18	16	19	30	21	24	30	25	31
At best tri-score combination	17	15	17	29	20	22	30	25	27
At tri-score combination using C2 tri-score wts.				30	21	24	31	25	29

sifications. Of the 134 subjects in population X′ receiving these tri-scores, a prediction as to which of the three diagnostic groups a person is *not* in would be incorrect in only 17, or 13%, of the cases. Had the diagnostic prediction of the 16 subjects not receiving any of the above tri-scores been made on a chance basis, the percentage incorrectly identified would still be under 17%.

6. *Comparison of A Priori Judgments as to Diagnostic Significance of Responses with Empirical Results*

The a priori judgments were compared to the empirical results in regard to the diagnostic implications (normal, anxiety reaction or schizophrenic) of the various percepts. The results of this comparison are presented in Table 8. Listed at the left are the diagnostic categorizations which were determined empirically. These categorizations were based on the results obtained with the entire experimental population. The letter "O" indicates those percepts which did not receive any weights (using the $p = .10$ weighting system), and hence could not be considered characteristic of either diagnostic group.

The a priori judgments as to diagnostic implication are presented along the top of the table. The "?" stands for the alternative "something else," while the letter "F" is the response "failure to select one of the alternatives in a question." Separate results were computed for Test AB and Test C. In the former, it was possible to determine the empirical weight for each percept exactly. In the latter, however, this determination had to be made qualitatively.

In Test C, the subjects checked one of five possible ratings for each percept, and a separate weight was found for each of these five ratings. For some percepts, the diagnostic implication was quite clear since the first few rating steps were characteristic of one diagnostic category, while the last few were characteristic of the other diagnostic group. However, the picture generally was more confused. For example, the first and fifth steps might be given a normal weight, the fourth step an anxiety reaction weight, and the second and third steps given no weight. With results such as these, it was impossible to define the percept as characteristic of normals or of anxiety reactions. In many cases, borderline decisions had to be made, such as when of the five rating steps only the first or second step was found to be weighted. Included in the present table were only those percepts whose diagnostic implications seemed fairly clear.

Table 8 reveals that for the N-AR comparison pair there is no correlation between a priori judgment and the empirical results.[1]

Of the 60 percepts considered to be characteristic of normals on Test AB, only 9 proved to be so empirically, and in four cases the supposed normal percepts were found to be characteristic of anxiety reactions. Of the 80 percepts judged to be characteristic of anxiety reactions, only 6 were so empirically, while 13 were found to be characteristic of normals. Thus, in only 15 of the 140 percepts did the a priori judgment coincide with the empirical evaluation, while in 17 cases they were exactly the opposite. Very similar results were also obtained on Test C.

---

[1] It should be clearly recognized that the authorities who estimated the psychopathological significance of the responses did so on the basis of their experience with the individual Rorschach, and they made no claims that the significance would remain the same on a multiple choice test.

For the other two comparison pairs, there appears to be a positive correlation between a priori judgment and empirical evaluation. In the N-S comparison, of the 60 percepts presumed to be normal on Test AB, 16 actually were so with only 5 being exactly the opposite; that is, being characteristic of schizophrenics. Of the 80 percepts estimated as schizophrenic, 29 were found to be so empirically with only 6 having exactly the opposite diagnostic significance. Thus, in 45 of the 140 percepts the a priori judgments and the empirical results coincided, while they were exactly the opposite for only 11 percepts. The results of Test C, though, are not quite as good, with only 31 of the percepts coinciding while 19 were exactly the opposite.

In the AR-S comparison it is Test C which shows the greater amount of correspondence between a priori judgment and empirical evaluation. Here 42 of the percepts coincide, with only 15 being exactly the opposite. However, little relationship is shown in Test AB where 29 percepts coincide, and 25 are exactly the opposite.

It is difficult to use a single statistic to indicate the amount of correspondence because a lack of coincidence between the a priori judgment and the empirical results may be expressed in two ways: the estimated percept may prove to be significant of neither diagnostic category; or, what appears to be more serious, it may turn out to be significant of exactly the opposite diagnostic category than was expected.

In general, it may be said that a positive correlation exists between a priori judgments and empirical results since these two coincide more frequently, on the whole, than they are exactly the opposite. However, this correlation seems to be fairly low as can be seen from the relatively few times a percept is found empirically to have exactly the same significance as was expected on an a priori basis. Thus, while a priori judgments may be of value in the initial selection of items for a multiple choice projective test, there can be no question but that it will be essential that the

TABLE 8

CORRESPONDENCE BETWEEN EMPIRICAL AND A PRIORI CATEGORIZATIONS

Comparison Pair	Empirical Categorization	A Priori Categorization								
		Test AB					Test C			
		N	A	S	?	F	N	A	S	F
N-AR	Normal	9	13	6	2	0	14	12	2	0
	Anxiety Reaction	4	6	19	4	1	4	5	16	5
	Neither	47	61	55	14	59				
N-S	Normal	16	19	6	4	0	17	8	6	0
	Schizophrenic	5	12	29	5	35	13	9	14	67
	Neither	39	49	45	11	25				
AR-S	Anxiety Reaction	14	15	10	2	0	18	19	1	0
	Schizophrenic	3	15	14	1	26	12	14	23	45
	Neither	43	50	56	17	34				

psychopathological significance of each percept be determined empirically.

However, the data in Table 8 also indicate that a detailed analysis of those percepts found to be significant of a diagnostic category may uncover general principles applying to multiple choice projective tests which could materially increase the validity of a priori judgments. For example, in the N-AR comparison, the a priori estimations would have been much more accurate had those percepts which were considered to be characteristic of schizophrenics been judged to be indicative of anxiety reactions, while those percepts predicted as typical of anxiety reactions were instead considered characteristic of normals. This holds true for both Test AB and Test C. Similarly, in the AR-S comparison, the percepts which were thought to be indicative of anxiety reactions proved to be much less discriminating than the percepts which has been assumed to be characteristic of normals. Also apparent from Table 8 is the fact that the alternative "Something else" (?) does not appear to differentiate any diagnostic group, while the "Failure to select an alternative" (F) seems to be primarily a schizophrenic response.

7. *Comparison of Conditions Varied within the Test Battery*

As outlined previously (Section II,B, 2), the experimental design included variations in certain of the test conditions. By comparing the results of different sections of the test battery, the effect of these variations could be determined. The relative discrimination obtained under the different test conditions is shown in Table 9. While the amount of discrimination under each condition was measured by both the critical ratio and the % Wrong (percentage of subjects incorrectly identified), only the critical ratios are listed because of the extremely high correlation between these two statistics. Preference was given to the critical ratio because it would seem to be the more reliable statistic.

a. *Differences in the structure of the questions.*

(1). *Relative vs. absolute decision by the subject.* In Test A and Test B,

TABLE 9

COMPARISON OF CONDITIONS VARIED WITHIN THE TEST BATTERY

Condition	Test Section	Critical Ratios for Different Comparison Pairs		
		N-AR	N-S	AR-S
Relative vs. absolute decision by subject	Test AB (relative)	3.4	5.5	4.0
	Test C (absolute)	3.0	7.8	4.1
Difference in type of percepts in a question	Design I (forced normal)	0.7	4.1	2.7
	Design II (unlimited)	3.8	3.7	2.8
	Design III (forced non-normal)	1.6	4.6	3.0
Brief vs. lengthy exposure of stimulus	Test A (brief)	2.6	4.1	3.5
	Test B (lengthy)	3.2	5.9	2.5
No. of questions in the test battery	Designs I, II, III (20 questions)	2.0	4.5	2.8
	Test A, B (30 questions)	2.9	5.0	3.0
	Test AB (60 questions)	3.4	5.5	4.0
	Test C (100 questions)	3.0	7.8	4.1
	Test ABC (160 questions)	3.9	7.1	4.8

which have been combined to form Test AB, the subjects were required to select one of four alternatives, thus making a relative decision among these alternatives. On Test C, the subjects rated each alternative on a five point absolute rating scale. As can be seen from the results in Table 9, the type of decision made by the subjects appears to have little effect upon the amount of discrimination. While Test AB is somewhat superior for the N-AR comparison pair, Test C has a higher critical ratio for the N-S comparison pair, while the results are practically identical in the AR-S comparison pair.

(2). *Differences in the types of alternatives included in a question.* Results were obtained for three different question "Designs," each of which was structured to include only certain types of percepts (section II,B,2a [2]). It was necessary to select percepts on the basis of an a priori judgment as to their psychopathological significance, and, as has been shown, such a priori judgments do not correspond to the psychopathological significance found empirically. Hence, the original rationale for this comparison no longer exists. However, there remains a comparison between questions in which the subjects are required to select a specific percept (Designs I and III), and questions in which the subject may avoid selecting a specific percept by checking the alternative, "Something else," (Design II).

From the results it is seen that Design II was superior for the N-AR comparison pair, while Design III was best for the N-S comparison pair. Little difference was found among the designs in the AR-S comparison pair. While no superiority seems to be shown for either the "forced-choice" or the "unlimited-choice" question designs, any conclusions should be made cautiously since these results were based on only 20 questions of each design.

b. *Differences between a brief and lengthy exposure of the stimulus.* In Test A the stimuli were exposed for a relatively brief period. In Test B, which was equivalent to Test A in other respects, the subjects were given as much time as they desired to view the stimulus. As shown in Table 9, Test B produced greater discrimination in the N-AR and N-S comparison pairs, while Test A was superior in the AR-S comparison pair. These results do not appear to be definitive enough to warrant drawing any general conclusions. It remains possible that the differences found represent true differences, and that opposite results are obtained in different comparison pairs because "optimum" conditions are different for each pair of diagnostic groups.

c. *The effect of increasing the length of the test.* By considering different sections of the test battery separately, it was possible to determine empirically the effect of increasing the length of the test. Designs I, II and III each included only 20 questions, so the mean of the critical ratios obtained with these "Designs" was used to estimate the discrimination produced by 20 questions. The amount of discrimination resulting from 30 questions was determined by the mean of the critical ratios obtained with Test A and Test B, each of which contained that number of questions. When Tests A and B were combined to form Test AB, there was a total of 60 questions. In Test C there were 100 questions, while the complete battery, Test ABC, contained 160 questions. The fact that the types of questions were different for different test lengths was not considered contaminating since it would appear from the preceding results that the type of question has little effect upon the amount of discrimination.

As can be seen in Table 9, there is, in general, a steady increase in discrimination for each comparison pair as the length of the test increases. Were the critical ratios of the three comparison pairs to be averaged, then the few fluctuations would be eliminated, and the amount of discrimination would show a regular progression. The importance of the length of the test upon individual prediction can be gathered from a comparison of the percentage of subjects incorrectly identified. The mean % Wrong of the three comparison pairs drops from 35% for the 20 question section to 25% for the 160 question battery.

8. *The Effect of Age and Education*

Since there were some differences among the three diagnostic groups with respect to age and amount of education, it was possible that the discrimination obtained between diagnostic groups was a reflection, in part at least, of these age and educational differences. In Table 10, the correctly and incorrectly identified subjects (defined in section V,B,1,b) in each comparison pair are differentiated with respect to age and amount of education. For purposes of comparison, data are also presented for all of popula-

TABLE 10

COMPARISON OF SUBJECTS INCORRECTLY IDENTIFIED AND SUBJECTS CORRECTLY IDENTIFIED, WITH RESPECT TO AGE AND AMOUNT OF EDUCATION

Diagnostic Category	Compar. Pair	Group	Age Comparison			Education Comparison		
			N	Mn	Mdn	N*	Mn	Mdn
Normals	N-AR	Incorrect	10	25.2	25.2	10	11.7	12.5
		Correct	40	25.8	24.3	40	12.5	12.7
	N-S	Incorrect	10	25.5	23.5	10	11.6	11.7
		Correct	40	25.7	25.1	40	12.5	12.8
	Both	Pop. X	50	25.6	24.7	50	12.3	12.7
		Total Pop.	250	25.0	23.8	245	13.2	12.8
Anxiety Reactions	N-AR	Incorrect	10	33.3	33.0	10	12.3	13.0
		Correct	40	30.1	29.5	40	10.6	10.6
	AR-S	Incorrect	10	35.1	36.0	10	8.8	8.8
		Correct	40	29.6	29.5	40	11.4	12.1
	Both	Pop. X	50	30.7	30.5	50	10.9	10.9
		Total Pop.	203	30.0	29.6	197	11.4	11.5
Schizophrenics	N-S	Incorrect	10	27.3	25.5	8	12.3	12.0
		Correct	40	29.2	28.5	34	11.0	11.6
	AR-S	Incorrect	10	26.6	25.5	9	11.8	11.5
		Correct	40	29.4	28.8	33	11.1	11.6
	Both	Pop. X	50	28.8	28.0	42	11.3	11.6
		Total Pop.	200	28.4	27.3	179	11.5	11.6

* For some subjects information as to the amount of education was not available.

tion X, which includes both correctly and incorrectly identified subjects, and for the entire experimental population.

Each diagnostic group is included in two comparison pairs. The results of the normals and schizophrenics are very similar in both comparison pairs because in each case many of the same subjects were incorrectly identified in both comparison pairs. Six of the normals who were incorrectly identified in the N-AR comparison pair were also incorrectly identified in the N-S comparison pair. With the schizophrenics, there were nine subjects duplicated. However, none of the anxiety reactions was incorrectly identified in both comparison pairs. (A chance matching of 10 names in each diagnostic category would have resulted in two subjects being included in both comparison pairs.)

There seems to be no difference in age between correctly and incorrectly identified normals. However, the incorrectly identified schizophrenics are a few years younger than those correctly identified, while the incorrectly identified anxiety reactions, particularly in the AR-S comparison pair, are an older group.

The incorrectly identified normals have slightly less education, while the incorrectly identified schizophrenics are slightly better educated. With the anxiety reactions, however, the educational differences are marked. Those incorrectly identified as normals have about two more years of education than those correctly identified, while the anxiety reactions identified as schizophrenics have about three years less education.

From these results it would appear that, particularly in the case of the anxiety reactions, the subjects' responses are a product of their educational background as well as of their psychopathological condition. The educational background may, in turn, be a reflection of intellectual endowment. The entire question appears to be too complex to be resolved in the present study, particularly since the inadequacy of the original diagnostic classification may also be a factor. The chief screening psychiatrist, in charge of the panel which diagnosed almost all the anxiety reactions in the cross validation population, stated that most diagnostic errors probably occurred when attempting to determine whether the older patients with little education (i.e., those anxiety reactions identified on the SMCR as schizophrenics) were anxiety reactions or schizophrenics. In any case, the overlap between correctly and incorrectly identified subjects with respect to age and education was so great that differences in these factors cannot be responsible for most of the discrimination obtained with the SMCR.

VI. DISCUSSION OF THE RESULTS

Upon cross validation, the SMCR incorrectly identified at most 30% of the N-AR comparison pair, 25% of the N-S comparison pair, and 30% of the AR-S comparison pair. For all comparison pairs the critical ratios were sufficiently high to establish a significant difference between diagnostic groups. By using tri-scores, the diagnostic category a subject was *not* in could be predicted with an error of under 17%. Therefore, it would appear that a fair validity had been obtained with a multiple choice projective test. Since the amount of discrimination found compares favorably with the validity reported for other screening measures, and since there are indications that subsequent research with multiple choice projective tests will lead to even greater discrimination, further investigation with this approach would seem warranted.

A. Comparison with Other Screening Devices

As mentioned previously, screening tests have been found to be most effective in military screening. Ellis and Conrad (10) have presented a detailed summary of the results obtained with personality inventories in this area. For most studies, the amount of validity reported is approximately the same as that obtained with the SMCR. While there were a few reports of unusually high validity, Ellis and Conrad suggest that such validity coefficients reflect the operation of unusual experimental or statistical factors.

The possible effect of improper cross validation, an experimental error found in many studies, has been well illustrated by data obtained in the present investigation. When a population of 50 subjects in each diagnostic group (C1) was scored with weights derived from the same population, the scores from just half the test battery resulted in only about 8% of the subjects being incorrectly identified. When weights derived from a criterion population of 150 subjects in each diagnostic category (C123) were used to score one-third of these subjects (C2 and C3), the percentage of subjects incorrectly identified rose to an average of 18.5%. When the weights derived from this criterion population were applied to a cross validation group, the percentage of subjects incorrectly identified at the best critical score of the cross validation group became 25%. When a critical score derived from the criterion population was employed, the average percentage of subjects incorrectly identified rose to 28%.

Zubin (60), in another article evaluating screening tests, presents the results obtained at the Sheepshead Bay U. S. Maritime Training Station as an example of the marked effectiveness of a questionnaire in military screening which generally has produced better results than civilian screening. Of 1,000 men processed, 660 were passed without a further interview on the basis of a screening personality questionnaire. Three of those passed had to be discharged from the service during their training period for psychiatric reasons. Of the 340 men selected for a psychiatric evaluation, 15 had to be discharged from the service, 75 had temporary or mild emotional conditions which required short term therapy, while the remaining 250 turned out to be false positives.

In evaluating the above data, the 75 subjects who adjusted after short term therapy are difficult to place. There may have been an equal proportion of subjects among the persons passed without a psychiatric interview who had comparably mild emotional conditions, but who were able to adjust without any short term therapy. If these 75 subjects were therefore to be considered as false positives, there would be a total of 1,000 subjects, 982 passing and 18 being discharged. Of these, 33% (325 passing and 3 discharged) were incorrectly identified. Eliminating the 75 subjects from the data would lower the percentage of subjects incorrectly identified to 27%.

Not only are the above percentages about the same as those obtained with the SMCR, but they probably give a spuriously high estimate of validity. Only persons making high questionnaire scores would have been discharged (on the basis of a psychiatric examination) without participating in the regular training program. Ellis and Conrad report a study by Weinstock and Watson "on 121 Naval recruits who were allowed to remain in service ('on trial') despite an adverse prognosis based on clinical judgment. Of the 121 recruits, only 44 (36%) were discharged for neuropsychiatric reasons during recruit training" (10, p. 386).

Another factor which frequently leads to spuriously high validity estimates is the use of extreme groups. With such populations almost any screening measure can obtain a fair validity. As Huston and Shakow found in a study on schizophrenics, while the patient groups performed at a level significantly poorer than the control group on two indices of steadiness, "when these patients who cooperated as well as the normal group were compared with the latter, the differences between the groups disappeared" (25, p. 120). It was for this reason that in the present study care was taken to test only the "best" schizophrenics, and to use anxiety reactions who were not so severely disturbed as to warrant hospitalization.

Had a typical institutionalized population been used, the SMCR would undoubtedly have produced a much greater amount of discrimination. For example, there were 48 schizophrenic subjects who filled out Tests A and B correctly according to the established criteria, but who failed to fill out Test C properly. The greater failure on Test C resulted from the fact that it required a somewhat more complex response and came last in the testing battery when increased boredom and fatigue had set in. When the scores of the 48 schizophrenics who failed to complete Test C were substituted for the regular schizophrenic cross validation population in the Test AB comparisons, the percentage of subjects incorrectly identified at the best critical score dropped from 24% to 10% in the N-S comparison, and from 32% to 22% in the AR-S comparison. Furthermore, had the SMCR been administered so that a higher proportion of schizophrenics would have responded, such as under individual conditions or with proctoral assistance, the percentage of subjects incorrectly identified would undoubtedly have been even smaller.

B. Indications for Greater Potential Discrimination

While there has been a tremendous amount of experimentation with questionnaires extending back to the first World War, relatively little work has been done with multiple choice projective tests. Indications exist from the present study that further research with the multiple choice projective approach may lead to a marked increase in the amount of discrimination produced.

Firstly, the present study did not attempt to exhaustively sift questions until a test composed of the most discriminatory choices was evolved. By far the greater number of responses received no weights, even under the most liberal weighting system. Yet the greater amount of discrimination occurring as the size of the test was increased clearly indicates that the replacement of non-discriminatory percepts should add even more to the validity.

Secondly, only one projective method, the Rorschach, was used. It is recognized in individual psychopathological testing that much more validity can be obtained from a battery of different projective tests than from any single instrument. As Rapaport and Schafer have stated: ". . . as an exclusive basis of diagnosis the test (the Rorschach) has serious limitations and its real potency is exploited only when used as the center of a battery of tests" (42, p. 73).

Thirdly, a loss in validity undoubtedly occurred because of inadequate diagnoses of subjects in the three diagnostic classifications. The Cornell Index, which was used as the criterion of normality is itself only a rough screening device, and it has not proved too effective in the screening of prepsychotic and early psychotic states. Norms for Form N1 of the Cornell Index could not be obtained. However, 10 or fewer poor responses on Form N2 which contains 37 additional questions and consequently provides more opportunity for poor responses, will be made by 14% of the persons rejected for military service for psychiatric reasons. Thus, it is probable that some subjects in the "normal" population were poorly adjusted. Furthermore, some of the schizophrenics were in remission, or at least well enough adjusted to be released from the hospital within a few weeks of their testing, while others had a great deal of free floating anxiety characteristic of anxiety reactions. The anxiety reaction group was perhaps the most poorly diagnosed, containing many individuals who would not have contemplated psychiatric treatment had they not been veterans, as well as persons whose anxiety was symptomatic of a latent schizophrenia. Almost all the anxiety reactions in the cross validation population were selected only on the basis of a screening or tentative psychiatric diagnosis.

Of course, most subjects in each diagnostic category were probably correctly classified. However, the number of subjects incorrectly identified by the test was relatively small, so that the inclusion of some misclassified persons in this group would make an appreciable difference. It is recognized that the misclassified subjects may not be the persons incorrectly identified by the test. However, the possibility appears to be sufficiently great to warrant a more thorough evaluation of the subjects in each diagnostic category before the maximum potentiality of the test can be estimated.[1]

---

[1] Since a year had elapsed between the testing of the subjects and the preparation of this monograph for publication, an attempt was made to check the accuracy of the diagnosis established at that time by a follow-up study. No further information could be obtained for the normal subjects since they had filled out the tests anonymously. However, psychiatric progress reports were available for most of the anxiety reactions and schizophrenics.

The majority of the anxiety reactions tested were subsequently treated by psychiatrists working in private clinics under contract to the Veterans Administration. These psychiatrists prepared monthly reports summarizing the condition of their patients. The majority of the schizophrenics had been tested at the Veterans Hospital, Northport, L. I., and for these subjects complete clinical case folders were available.

Unfortunately, a careful inspection of this data was not helpful in improving the validity of the original diagnosis. While almost all of the subjects retained their initial diagnosis, a consideration of their symptomatology revealed much overlapping between subjects diagnosed as anxiety reactions and those diagnosed as schizophrenics. It appeared that psychiatrists at out-patient clinics generally preferred to make a diagnosis of anxiety reaction while those at neuropsychiatric hospitals were more likely to make a diagnosis of schizophrenia. This is in agreement with the experience of the writer at both types of installations, and with the opinions of several psychiatrists consulted on this question. Thus, subjects with identical symptoma-

Finally, a multiple choice projective test combined with other types of tests in a battery may result in a greater amount of discrimination than would exist for any of the tests separately. For example, a test such as the SMCR which resulted in greater discrimination between normals and psychotics might be combined with a questionnaire such as the Cornell Index which is more effective in differentiating normals from neurotics. Subjects might also be given an intelligence test, and their multiple choice projective test responses interpreted according to their I.Q. It is recognized that responses on the free association Rorschach will depend upon a subject's intelligence as well as his psychopathological condition. A "vague" or "mediocre" response which might be considered normal for a subject of low intelligence would be indicative of maladjustment in one of high intelligence. The same would appear to be true of the multiple choice Rorschach. In the present investigation, the large difference in education between the anxiety reactions identified incorrectly as normals and those identified incorrectly as schizophrenics probably reflects a difference in intelligence rather than in education per se. Therefore, it might be advisable to develop different sets of scoring weights for different levels of intelligence.

C. SUGGESTIONS AS TO FUTURE RESEARCH

From the results of the present study, both quantitative and qualitative, a number of suggestions arise as to the course of future investigations.

1. *Subjects*

In the present study, weights based on the results of only 50 subjects in each diagnostic category proved inadequate. On the other hand, a criterion population of 150 subjects per diagnostic category was no more effective than one of 100 subjects. But caution would be necessary in applying these results to other investigations. For example, a smaller number of subjects might be required when only two or three responses are possible for each question rather than the 5 or 6 in the SMCR, or when a psychopathological variable less heterogeneous than diagnostic categories is investigated. However, the lack of very marked differences between diagnostic groups in their responses to any question in the SMCR indicates that a fairly large number of subjects will probably be required to establish reliable scoring weights.

Not only should the subjects in both criterion and cross validation populations be very competently diagnosed, but it would be valuable to have them evaluated psychiatrically on a wide variety of variables. In addition to nosological classification, the subjects could be rated upon significant personality traits, such as hostility or dependence, and upon dynamic factors such as sibling rivalry or parental rejection. Then, the same subject population could be regrouped accordingly when studying each different psychopathological variable. A prognostic index could also be developed by reconsidering the data after the progress of the subjects had been observed.

2. *Stimuli*

Further use of the Rorschach inkblots as stimuli in preference to other inkblots would seem inadvisable. As the present study has shown, there is little correlation with respect to the psychopathological significance of responses between a priori judgments based on the free association Rorschach and the results of the multiple choice Rorschach. Therefore, the large experiential background which has been built up with these particular inkblots under free association test conditions would be of little value in the construction of a multiple choice test. As a disadvantage, the multiple choice Rorschach provides suggestions to the subjects by listing possible responses which may influence their performance on a subsequent free association Rorschach.[2]

---

tology might be diagnosed differently at different installations.

A further attempt to validate the SMCR on schizophrenics by comparing the scores of subjects who had been discharged with those who still remained in the hospital was abandoned when it was found that in a population of the "better" schizophrenics, discharge was determined as much by the patient's home environment as by his psychiatric condition.

[2] For this reason, the scoring stencils for the SMCR have not been included in the present report. It was felt that the availability of these

While no quantitative evidence is available, it was the experience of the writer and of several other clinical psychologists who tested subjects individually subsequent to the SMCR that the performance of these subjects had been influenced by the preceding multiple choice Rorschach. The evaluation of the subjects' free association protocols was hindered by the difficulty in assessing the direction and extent of the influence of the multiple choice test.

Another advantage in using other than the Rorschach inkblots is the opportunity to construct inkblots particularly suited for a multiple choice test. Such blots could be structured to emphasize responses to either large areas or the blot as a whole. In the present experiment, the subjects rarely studied the slides minutely, but based their responses on a broad general impression. (Of course, it is possible that a more detailed examination of the stimuli would have been made had the subjects been more highly motivated or had the inkblots been presented in booklet form.) The inkblots could be structured to emphasize certain determinants, such as Levy's graduated "movement" series (61), or they might present a choice between determinants such as by containing two possible percepts, one form determined and the other predominantly color determined. It may be desirable to draw inkblot designs with printer's ink, as was done by Terman and Miles (52), rather than to use actual inblots. While it would be desirable to explore the potentialities of indefinite figures placed on simple backgrounds, it is very possible that better results are obtained when both figure and ground are indefinite, as with inkblots. This reservation should be kept in mind for all types of projective stimuli.

In addition to inkblots, various other projective stimuli should be investigated. Here again, it would not appear advisable to use the same stimuli as in the free association tests, but rather to develop new stimuli which together with the multiple choice responses could be structured in the direction of a specific psychopathological variable. For example, rather than employing the TAT cards, special pictures could be drawn. To investigate the maternal relationship, a picture of a mother and a child, the former with her arm upraised, might be shown. The responses could include one in which the mother is petting the child, another in which she is about to strike the child, and a third in which she is waving to someone behind the child. Of course, the picture would have to be sufficiently vague so that the subjects' responses would depend upon their projections. With the Rosenzweig Picture Frustration Test (43, 44) which is already structured to measure a specific variable, that of frustration tolerance, it might be necessary only to develop a set of multiple choice responses for each of the pictures.

Geometric figures, such as those employed in the Bender Gestalt (1) and the Franck Symbolic Sex Preferences (11) could be used. Here, the alternatives might consist of other geometric figures, with the subjects required to select the figure most closely resembling the stimulus. Perhaps particularly well suited for the detection of paranoid tendencies would be the Tautophone (13) in which a recording of low-voiced nonsense syllables is played. Subjects are asked what the recording is saying, being given to understand that it is producing a meaningful statement.

Further work can be done with story plots along the lines already developed by Kelly, and Bishop, Taylor and Wolfenstein (27, 50, 59). Verbal stimuli could also be confined to single words in a variation of the word association approach. Either fragmented letters or very brief exposures during which only one or two letters would stand out might be employed, with subjects being asked which word had been displayed.

3. *Administration of the Test*

It would seem desirable to present the stimuli as tests of ability, and in fact, to include some questions with obviously correct responses to further this belief. The best exposure time would have to be worked out separately for each different stimulus. While the results of the present experiment indicated that the length of exposure of the Rorschach slides made little difference in the amount of discrimination obtained, no conclusion can be reached as to whether this will hold true for other stimuli or for other psychopathological variables which may be investigated.

It may be possible to diminish the effect of suggestion and rationalization in the selection of responses and thereby obtain more significant projections by removing the stimulus before the multiple choices are presented to the subjects. For example, in the SMCR this could be done by listing the percepts in a question on a second slide which would be projected on the screen after the stimulus had been removed. The subjects could indicate their choice of response by marking the appropriate column on an IBM answer form. With this technique, the subjects

---

stimuli might lead to an indiscriminate use of the test, thus exposing many persons to suggested responses with the Rorschach blots. The scoring stencils will, of course, be made available to any qualified psychologist who wishes to perform a confirmatory or further exploratory experiment.

might be forced to reply more upon their initial free associative responses to the stimuli.

4. *Structure of the Questions*

No definitive conclusions could be reached from the results of the present study regarding the structure of the questions. Similar results were obtained when the subjects made a relative decision by selecting one of four percepts and when they made an absolute decision by rating a percept on an absolute five point rating scale. Differences due to variations in question "design" (that is, the types of percepts contained in a question) could not be investigated as originally planned because the psychopathological significance of the percepts had been determined on an a priori basis, and there proved to be little correlation between a priori judgment and the empirical results. However, little difference was found between "forced choice" questions in which subjects had to select a specific percept, and "unlimited choice" questions in which they could check the alternative "Something else."

Because of the lack of relationship between a priori judgment and empirical findings, it would seem advisable in subsequent experimentation to determine first, insofar as is possible, the psychopathological significance of each response empirically from a simple three or four point rating scale. Then, after the significance has been established empirically, the percepts could be fitted into the structures of different question designs. These designs could extend from the relatively simple one suggested by Malamud and Malamud (32) in which the subjects select either one of two percepts listed or the alternative "neither," to a design with as many as 10 percepts, as in the Harrower Multiple Choice Rorschach (19).

5. *Length of Test*

It would seem desirable to have the test as long as possible. The present experiment demonstrated that the amount of discrimination increased with the length of the test up to at least 160 questions, and it may very well continue beyond that point. With the use of a variety of stimuli rather than a repetition of the same 10 inkblots, the subjects should not become bored or fatigued as quickly. Moreover, it was felt that the discriminating power of a stimulus decreased as additional questions were applied to that stimulus. Therefore, using a higher proportion of stimuli to questions than in the SMCR may also serve to increase the validity.

6. *Scoring of the Test*

Since the five weighting systems employed in the present investigation all produced about the same amount of discrimination, it would not seem to make much difference which weighting system were to be used. However, in addition to the simple weighting of items, it may be advisable to make a Guttman scale analysis (16) of the responses. Few psychopathological variables are "pure" entities. This is particularly true of nosological categories. For example, psychiatrists now prefer to speak of the "schizophrenias" rather than of "schizophrenia," believing that this diagnosis encompasses a variety of only loosely associated conditions. Similarly, many divergent psychopathological factors may lead to the symptomatic diagnosis of anxiety reaction. A scale analysis may disclose a relatively small combination of responses which are pathognomonic of a particular type of psychopathology but which do not materially affect the subjects' total score when all questions are weighted equally.

VII. CONCLUSIONS

I. Further exploration with the multiple choice projective test approach in the detection of psychopathology would seem warranted since:

A. A fair validity has ben obtained with a multiple choice projective test. When scoring-weights based on a sample of 100 subjects in each diagnostic category were cross validated on a new sample of 50 subjects in each category, the percentage of subjects incorrectly identified by the SMCR was at most 30% for the normal-anxiety reaction comparison, 25% for the normal-schizophrenic comparison, and 30% for the anxiety reaction-schizophrenic comparison. There was an error of under 17% in predicting which of the three diagnostic categories a subject would not be in.

B. Split half reliability coefficients, when corrected by the Spearman-Brown prophecy formula, were .77 in the normal-anxiety reaction comparison, .87 in the normal-schizophrenic comparison, and .72 in the anxiety reaction-schizophrenic comparison.

C. There are grounds for believing that even greater validity may be obtained by:

(1). Further experimentation with various multiple choice projection tests.

(2). Combining a multiple choice projective test with a questionnaire and an intelligence test in a screening battery.

II. Of the different test conditions varied in the present experiment, only the length of the test was shown to be significantly related to the amount of discrimination obtained.

III. No significant differences in the amount of discrimination could be established from:

A. The application of different methods of weighting responses.

B. Increasing the size of the criterion population upon which the weights were developed from 100 to 150 subjects in each diagnostic category.

C. The subjects responding by making a "relative" as opposed to an "absolute" decision.

D. The subjects being forced to choose among specific percepts in a question as opposed to being given the "unlimited" alternative, "Something else."

E. Exposing the stimuli for a brief period as opposed to a lengthy period.

IV. There was little correlation between a priori judgments based on the free association Rorschach and the empirical results obtained with the SMCR with respect to the psychopathological significance of responses.

V. Both age and education, the latter perhaps a reflection of intelligence, may affect the subjects' responses.

VIII. APPENDIX

A. THE SINGER MULTIPLE CHOICE RORSCHACH

Test A

NAME______________________________ AGE________ GRADE______ DATE__________

— 1 —
______A wolf's head
______An emblem
______A topographical map
______An octopus

— 2 —
______Ruins of a castle
______A blot
______A squashed bug
______An X-ray picture

— 3 —
______An animal split open
______Line down the center
______A butterfly
______Something else

— 4 —
______A bomb bursting
______Lungs and blood
______Two dogs
______Something else

— 5 —
______Animal heads
______Torn up butterfly
______A map
______Bears' heads

— 6 —
______Two equal sides
______Black and red paint
______Dark clouds
______A large eye

— 7 —
______Two red roosters
______Ribs
______A fish
______Lungs and chest

— 8 —
______Vertebra
______X-ray of the hips
______Person tearing someone apart
______Blood and black

— 9 —
______Smoke
______A hair ribbon
______Nothing at all
______Something else

— 10 —
______Flag of a country
______Mountain range
______A black waterhole
______Spilled ink

— 11 —
______A cow with fur
______Two feet
______An X-ray
______Something else

— 12 —
______Necks of geese
______Two scottie dogs
______A spinal column
______Sex organs

— 13 —
______A pair of antlers
______A butterfly with torn wings
______Dried blood
______Something else

— 14 —
______Legs
______A horrible object
______A map
______A butterfly

— 15 —
______A line down the center
______A bug
______A black mess
______Nothing at all

— 16 —
______A spattered mess
______A totem pole
______A backbone
______A turtle

— 17 —
______An oil explosion
______A bad design
______A cut up male sex organ
______Black paint

— 18 —
______An insect
______Jack the Ripper
______A riverbed
______Something else

— 19 —
______Intestines
______Ice formation
______A squashed frog
______A topographical map

— 20 —
______Two heads
______A coastline
______Nothing at all
______Something else

— 21 —
______A butterfly
______Grass
______A clown's face
______People watching each other

— 22 —
______Two bears climbing
______A biology slide
______A sailboat
______Something else

— 23 —
______Head of small dog
______Two polar bears climbing
______An Easter egg
______Modern art

— 24 —
______Colored butterflies
______An X-ray picture
______An airplane
______A blotch of colors

– 25 –

______Flowers
______Colored clouds
______Birth of a child
______Sea horses

– 26 –

______A dirty mess
______Smoke rising
______A bridge in the distance
______Nothing at all

– 27 –

______A person cut up
______A spinal column
______A halloween mask
______Something else

– 28 –

______Male sex organs
______Leaves
______Bugs
______People hanging in air

– 29 –

______Red and blue and yellow
______Bones
______Two spiders
______Something else

– 30 –

______A blue feathery flower
______A terrible germ
______Spiders attacking somebody
______Deer running

Test B

NAME______________________ AGE________ GRADE________ DATE________

– 1A –

______A bat
______Female sex organs
______A staring mask
______Something else

– 2A –

______Burnt piece of paper
______A spine
______A cat's face
______Two people

– 3A –

______A coastline
______A dark cloud
______A leaf
______Death

– 4A –

______Fire
______Two people
______A pair of pliers
______Smoke

– 5A –

______Human chest with a hole in it
______A squashed bug
______Blood
______A coastline

– 6A –

______Entrance to a tunnel
______Nothing at all
______A white lamp
______Something else

– 7A –

______Sex organs
______A skeleton effect
______Two staring eyes
______Picture from a medical book

– 8A –

______Waiters bowing
______A symmetrical design
______Animal cut open
______Something else

– 9A –

______Two devils
______Two monkeys hanging down
______A spinal cord
______Birds' heads

– 10A –

______A giant walking towards you
______Two snakes
______A furry man
______Something else

– 11A –

______A dragon skin
______A man sitting down
______A gorilla
______Snakes

– 12A –

______A relief map
______Clouds
______A torn chest
______Night

– 13A –

______Head of a mouse
______Profile of a man
______Antennae
______A dancer

– 14A –

______Two buffaloes crushing a rabbit
______Two equal sides
______Night
______Islands

– 15A –

______Dark smoke
______A bat
______The nervous system
______Something else

– 16A –

______A cat
______An X-ray
______A squashed worm
______Water

– 17A –

______Nothing at all
______Picture from an airplane
______Flying geese
______Something else

– 18A –

______Head of a dog
______A dragon fly
______A cloud
______A furry snake

– 19A –

______A Pacific island
______Two women go siping
______Male sex organs
______Something else

– 20A –

______A rabbit's head
______Clouds
______Scottie dogs
______Small dots on an island

– 21A –

______Smoke
______Just an inkblot
______A large eye
______Fog or mist

– 22A –

______A spinal column
______Decorated Christmas tree
______Flowers
______Ribs

– 23A –

______Pink, blue, orange
______Snow covered mountain
______Male sex organs
______A mountain

– 24A –

______Two fish
______Illustration in a medical book
______A coat of arms
______Something else

– 25A –

______A famous painting
______Staring eyes
______Red, green, orange
______Drawing in a medical book

– 26A –

______The liver
______Blood
______Two women
______Something else

– 27A –

______Horns of a deer
______An explosion
______Intestines
______A man's face

– 28A –

______Squashed insects
______A wishbone
______Snakes eating a rabbit's head
______Something else

– 29A –

______A tobacco pipe
______Two spiders
______Deer running
______Mountain sides

– 30A –

______A Christmas tree
______Modern art
______Bloody insides
______Illustration of different flowers

Test C

NAME______________________ AGE__________ DATE______________

	EXACTLY LIKE IT	VERY MUCH LIKE IT	SOMEWHAT LIKE IT	ONLY A LITTLE LIKE IT	NOT AT ALL LIKE IT
– I –					
1. A pelvis					
2. Two Santa Clauses					
3. Nothing at all					
4. A bat					
5. Female sex organs					
6. Woman with her arms raised					
7. A staring mask					
8. Small islands					
9. A squashed bug					
10. Ruins of a castle					
– II –					
11. A man with his head cracked open					
12. A butterfly					
13. Two clowns playing pattycake					
14. Nothing at all					
15. Entrance to a tunnel					
16. Faces on the edges					
17. Two equal sides					
18. An arrowhead					
19. A large eye					
20. A white lamp in the center					
– III –					
21. A person tearing someone else apart					
22. A bloody mess					
23. A coastline					
24. A symmetrical design					
25. Waiters bowing					
26. Men picking something up					
27. A bow tie					

	EXACTLY LIKE IT	VERY MUCH LIKE IT	SOMEWHAT LIKE IT	ONLY A LITTLE LIKE IT	NOT AT ALL LIKE IT
28. A butterfly					
29. Vertebra					
30. Nothing at all					
– IV –					
31. Head of an animal					
32. A furry man					
33. Two snakes					
34. A skin stretched out tight					
35. A mountain range					
36. A black waterhole					
37. Nothing at all					
38. A fur rug					
39. A giant walking towards you					
40. Two boots					
– V –					
41. The nervous system					
42. A line down the center					
43. A rabbit					
44. A smashed body					
45. Dark smoke					
46. A bat					
47. An insect					
48. A mountain range					
49. Nothing at all					
50. A nutcracker					
– VI –					
51. Nothing at all					
52. Leg of a table					
53. A snake with feathers					
54. Picture from an airplane					
55. Black paint					
56. A cat's whiskers					
57. Flying geese					
58. A butterfly					
59. An oil explosion					
60. Rug of animal skin					
– VII –					
61. Two women gossiping					
62. A topographical map					
63. Grass					
64. Male sex organs					
65. A white lamp					
66. Two female faces					
67. A Pacific island					
68. Nothing at all					
69. A necklace					
70. Rocks balancing					
– VIII –					
71. Animals climbing					
72. A coat of arms					
73. Nothing at all					
74. Two fish					

	EXACTLY LIKE IT	VERY MUCH LIKE IT	SOMEWHAT LIKE IT	ONLY A LITTLE LIKE IT	NOT AT ALL LIKE IT
75. Blood					
76. A blotch of colors					
77. An evergreen tree					
78. Frogs' heads					
79. Colored butterflies					
80. Modern art					
– IX –					
81. Nothing at all					
82. The liver					
83. A burning candle					
84. Blood					
85. Two men in dunce caps					
86. Two women					
87. Smoke rising					
88. Bloody hands					
89. A colored map					
90. Lobsters					
– X –					
91. A biology slide					
92. Leaves					
93. A wishbone					
94. Nothing at all					
95. Snakes eating a rabbit's head					
96. Flowers					
97. Two lions					
98. Two people hanging in air					
99. A rabbit's head					
100. Squashed insects					

B. Six-Item Questionnaire to Screen Normals

Answer each question YES or NO. If you are not sure, guess.

1. Have you ever had treatment for a nervous condition?
2. In general, do you consider yourself a nervous person?
3. While in the service, were you ever hospitalized for psychoneurosis, combat fatigue, nervousness or some similar condition?
4. Since you have gotten out of the service, have you generally (not merely occasionally) felt irritable or jumpy?
5. Do you get upset more easily than the average person?
6. Have you ever filed a disability claim or collected a pension for psychoneurosis?

IX. BIBLIOGRAPHY

1. Bender, Lauretta. A visual motor Gestalt test and its clinical use. *Res. Monogr. Amer. orthopsychiat. Assn.*, 1938, No. 3.
2. Blair, G. M. & Hoehn, A. J. The use of the Multiple Choice Rorschach test with adolescents, *J. educ. Res.*, 1947, **41**, 297-304.
3. Brody, M. B. Mental Testing. *J. ment. Sci.*, 1944, **90**, 127-151.
4. Chalke, F. R. C. The Harrower Stress Tolerance Test. *Psychosom. Med.*, 1946, **8**, 215-216.
5. Clark, R. M. A method of administering and evaluating the Thematic Apperception Test in group situations. *Genet. Psychol. Monogr.*, 1944, **30**, 3-55.
6. Crown, S. A Controlled Association Test as a measure of neuroticism. *J. Personality*, 1947, **16**, 198-208.
7. Deri, Susan K. *Introduction to the Szondi Test*, New York: Grune and Stratton: 1949.
8. Dorkey, Margaret, & Amen, Elizabeth W. A continuation study of anxiety reactions in young children by means of a projective technique. *Genet. Psychol. Monogr.*, 1947, **35**, 139-186.
9. Due, F. O., Wright, M. E., & Wright, B. A. The Multiple Choice Rorschach test in military psychiatric differentiation: The use of statistical criteria. (In Reference 19, *Large Scale Rorschach Techniques and Multiple Choice Test*. Springfield, Ill., Charles C Thomas, 1945, Section V).
10. Ellis, A., & Conrad, H. S. The validity of personality inventories in military practice. *Psychol. Bull.*, 1948, **45**, 385-426.
11. Franck, K. Preferences for sex symbols and their personality correlates. *Genet. Psychol. Monogr.*, 1946, **33**, 73,-123.
12. Frank, L. K. The projective methods for the study of personality. *J. Psychol.*, 1939, **8**, 389-413.
13. Grings, W. W. The verbal summator technique and abnormal mental states. *J. abnorm. soc. Psychol.*, 1942, **37**, 529-545.
14. Guilford, J. P. *Fundamental statistics in psychology and education*. New York: McGraw-Hill Book Co., 1942.
15. Guilford, J. P. Some lessons from aviation psychology. *Amer. Psychol.*, 1948, **3**, 3-11.
16. Guttman, L. A basis for scaling qualitative data. *Amer. Sociol. Rev.*, 1944, **9**, 139 150.
17. Harrower, Mollie R., & Grinker, R. R. The Stress Tolerance Test. Preliminary experiments with a new projective technique. *Psychosom. Med.*, 1946, **8**, 3-15.
18. Harrower-Erickson, Mollie R. Large scale investigations with the Rorschach. *J. Consult. Psychol.*, 1943, **7**, 120-127.
19. Harrower-Erickson, Mollie R., & Steiner, M. E. *Large scale Rorschach techniques and Multiple Choice Test*. Springfield, Ill., Charles C Thomas, 1945.
20. Harrower-Erickson, Mollie R., & Steiner, M. E. Modification of the Rorschach method for use as a group test. *J. Genet. Psychol.*, 1943, **62**, 119-133.
21. Hertz, Marguerite, Ellis, A., & Symonds, P. M. Rorschach Methods and other projective techniques. *Rev. educ. Res.*, 1947, **17**, 78-100.
22. Hertzman, M. A comparison of the individual and group Rorschach tests. *Ror. Res. Exch.*, 1942, **6**, 89-108.
23. Horowicz, Ruth E. A pictorial method for the study of self-identification in preschool children. *J. Genet. Psychol.*, 1943, **62**, 135-148.
24. Hunt, W. D., & Stevenson, I. Psychological testing in military clinical psychology. II Personality testing. *Psychol. Rev.*, 1946, **53**, 107-115.
25. Huston, P. E., & Shakow, D. Studies of motor function in schizophrenia. III Steadiness. *J. Gen. Psychol.*, 1946, **34**, 119-126.
26. Jensen, M. B., & Rotter, J. B. The value of thirteen psychological tests in officer candidate screening. *J. Appl. Psychol.*, 1947, **31**, 312-322.
27. Kelly, G. A. & Bishop, F. A projective method of personality investigation. *Psychol. Bull.*, 1942, **39**, 599.
28. Lawshe, C. H., Jr., & Forster, M. H. Studies in projective techniques: I The reliability of a multiple choice group Rorschach test. *J. Appl. Psychol.*, 1947, **31**, 199-211.
29. Lindner, R. M. A further contribution to the group Rorschach. *Ror. Res. Exch.*, 1943, **7**, 2-15.
30. Lindner, R. M., Chapman, K. W., & Rinck, E. C. The development of a group Rorschach technique in a federal penal institution with special reference to the problem of psychopathic personality. *Psychol., Bull.*, 1942, **39**, 513-514.
31. Malamud, D. I. Value of the Maller Controlled Association Test as a screening device. *J. Psychol.*, 1946, **21**, 37-43.
32. Malamud, Rachel F., & Malamud, D. I. Validity of the amplified multiple choice Rorschach as a screening device. *J. Consult. Psychol.*, 1945, **9**, 224-227.
33. Malamud, Rachel F., & Malamud, D. I. The multiple choice Rorschach: a critical examination of its scoring system. *J. Psychol.*, 1946, **21**, 237-242.
34. Maller, J. B. *Controlled Association Test*

(Forms A and B). New York: Bureau of Publications, Teachers College, Columbia Univ. 1934.

35. Michael, J. C., & Buhler, C. Experiences with personality testing in a neuropsychiatric department of a public general hospital. *Dis. nerv. Syst.*, 1945, **6**, 205-211.
36. Mittlemann, B., & Brodman, K. The Cornell Indices and the Cornell Word Form I. Construction and Standardization *Ann. N.Y. Acad. Sci.*, 1946, **46**, 573-577.
37. Munroe, Ruth. Prediction of the adjustment and academic performance of college students by a modification of the Rorschach method. *Appl. Psychol. Monogr.*, Stanford Univ. Press, 1945, **7**, pp. 104.
38. Murray, H. A., & Stein, M. A note on the selection of combat officers. *Psychosom. Med.*, 1943, **5**, 386-391.
39. Pressey, S. L. A group scale for investigating the emotions. *J. abnorm. soc. Psychol.*, 1921, **16**, 55-64.
40. Pressey, S. L., & Pressey, L. W. "Cross-Out" tests with suggestions as to a group scale of the emotions. *J. Appl. Psychol.*, 1919, **3**, 138-150.
41. Rapaport, D. The Szondi Test. *Bull. Menninger Clin.*, 1941, **5**, 33-39.
42. Rapaport, D., & Schafer, R. The Rorschach test: a clinical evaluation. *Bull. Menninger Clin.*, 1945, **9**, 73-77.
43. Rosenzweig, S. The picture association method and its application in a study of reactions to frustration. *J. Personality*, 1945, **14**, 3-23.
44. Rosenzweig, S., Clarke, H. J., Garfield, M. S., & Lehndorff, A. Scoring samples for the Rosenzweig Picture-Frustration Test. *J. Psychol.*, 1946, **21**, 45-72.
45. Ross, W. D., & Ross, S. Some Rorschach ratings of clinical value. *Ror. Res. Exch.*, 1944, **8**, 1-9.
46. Sargent, Helen. Projective Methods: their origins, theory and application in personality research. *Psychol. Bull.*, 1945, **42**, 257-293.
47. Springer, N. N. The validity of the multiple choice Rorschach in the screening of naval personnel. *J. Gen. Peychol.*, 1946, **35**, 27-32.
48. Symonds, P. M., & Krugman, M. Projective methods in the study of personality. *Rev. educ. Res.*, 1944, **14**, 81-98.
49. Symonds, P. M., & Samuel, E. A. Projective methods in the study of personality. *Rev. educ. Res.*, 1941, **11**, 80-93.
50. Taylor, W. S. Reality and fantasy thinking in social reconstruction. *Indian J. Soc. work.*, 1946, **7**, 169-175 (abstract).
51. Temple, Rita, & Amen, Elizabeth W. A study of anxiety in young children by means of a projective technique. *Genet. Psychol. Monogr.*, 1944, **30**, 59-113.
52. Terman, L. M., & Miles, Catherine C. *Sex and personality (Studies in masculinity and femininity).* New York: McGraw-Hill Book Co., 1936,
53. Weider, A., Brodman, K., Mittlemann, B., Wechsler, D., & Wolff, H. G. The Cornell Index: a method for quickly assaying personality and psychosomatic disturbances to be used as an adjunct to interview. *Psychosom. Med.*, 1946, **8**, 411-413.
54. Weider, A., Mittlemann, B., Wechsler, D., & Wolff, H. G. The Cornell Selective Index. *J. Amer. Med. Assn.*, 1944, **124**, 224-228.
55. Weider, A., & Wechsler, D. The Cornell Indices and the Cornell Word Form. 2. Results. *Ann. N.Y. Acad. Sci.*, 1946, **46**, 579-591.
56. Weider, A. *et al. Cornell Index Manual.* New York: The Psychological Corporation, 1946.
57. White, R. W. Interpretation of Imaginative Productions. (From Hunt, J. McV. *Personality and the behavior disorders,* New York: Ronald Press, 1944. Vol. I, Chapt. VI.)
58. Wittman, P. The multiple choice Rorschach test as a differential diagnostic tool. *Arch. Neurol. Psychiat.* Chicago, 1946, **55**, 154-156.
59. Wolfenstein, M. The reality principles in story preferences of neurotics and psychotics. *Character & pers.*, 1944, **13**, 135-151.
60. Zubin, J. Recent advances in screening the emotionally maladjusted. *J. Clin. Psychol.*, 1948, **4**, 56-63.
61. Zubin, J., Levy, D. M., & Rust, R. Movement responses in normals, schizophrenics and neurotics on the Levy Movement Cards. *Amer. Psychol.*, 1947, **2**, 269.

VOLUME 64
NUMBER 9

WHOLE NO. 315
1950

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

A Normative Study of the Thematic Apperception Test

By

LEONARD D. ERON
Yale University

Accepted for publication, November 28, 1949

Price $1.50

Published by

THE AMERICAN PSYCHOLOGICAL ASSOCIATION
1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

COPYRIGHT, 1950, BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION

ACKNOWLEDGMENT

IT IS A PRIVILEGE to be able to thank the many individuals who aided in this study. Dr. Ann Magaret, University of Wisconsin, shared in the conception of the problem and in establishing the criteria for analysis. Dr. Norman Cameron, University of Wisconsin, contributed extremely worthwhile advice with regard to interpretation of findings and clarity of presentation. Dr. David A. Grant, University of Wisconsin, helped with the experimental design and the appropriate use of statistical methods. Mrs. Dorothy T. Hake and Mr. Robert Callahan of the University of Wisconsin collaborated with me in the construction of rating scales of emotional tone and outcome and Mrs. Hake rated all 3,000 stories in this series on these variables. Drs. Irvin L. Child, Carl I. Hovland and Seymour B. Sarason of Yale University contributed helpful suggestions. Dr. Herbert S. Conrad, Editor, and Dr. Eugenia Hanfmann, Consulting Editor of Psychological Monographs, made important recommendations as to presentation of results and interpretation. To all these persons and to the 150 subjects, who of course must remain anonymous, I wish to express my sincerest gratitude for their cooperation and interest.

LEONARD D. ERON

TABLE OF CONTENTS

ACKNOWLEDGMENT iii

I. INTRODUCTION 1

II. PROCEDURE 2

A. Subjects 2

B. Material and Administration 3

C. Technique of Analyzing Stories 4

1. Thematic Analysis 4

2. Identification of Characters 4

3. Level of Interpretation 5

4. Emotional Tone 5

D. Statistical Method 5

III. RESULTS AND DISCUSSION 7

A. Emotional Tone of Stories 7

B. Outcome of Stories 8

C. Shift in Tone 9

D. Perceptual Distortions 11

1. Misidentification of Sex 11

2. Other Perceptual Distortions 12

E. Themes 13

F. Level of Interpretation 22

IV. NORMATIVE DATA FOR THE TAT 25

V. SUMMARY AND CONCLUSIONS 28

APPENDIX A: RATING SCALE OF EMOTIONAL TONE 32

APPENDIX B: CHECK LIST OF THEMES 34

APPENDIX C: CARD BY CARD NORMATIVE MATERIAL 37

BIBLIOGRAPHY 47

A NORMATIVE STUDY OF THE THEMATIC APPERCEPTION TEST[1]

I. INTRODUCTION

A NUMBER of studies have indicated that the Thematic Apperception Test is a valid instrument for measuring the usual product of fantasy central to an individual (25, 36, 40, 49), and thus many investigators have assumed that this technique can be used as a basis for diagnostic differentiation among clinical groups. Preliminary investigations by this writer, however, have indicated that no broad group differences could be elicited, at least in respect to content (16) and emotional tone (19), and that hospitalization itself might be an important influence in determining the nature of the productions (19). It was also disclosed in these studies that many of the diagnostic interpretations placed on certain characteristics of TAT stories did not apply to the groups of subjects they included and that there thus was a definite need for normative data comparable with the Rorschach tables of Beck (6) and Hertz (27). Were such data available to all who use the TAT, many false impressions about certain records would be corrected; and there would be less danger than there has been of the examiner's projecting his own biases into personality interpretations constructed on the basis of the TAT. Furthermore, the dangerous practice of "cue" analysis,[2] often based on faulty analogical reasoning (17), and susceptible to misuse by the inexperienced, might be considerably curtailed if the examiner had recourse to adequate normative data.

It is the purpose of the present research to clarify three fundamental questions which were raised by these preliminary investigations:

1. Do the fantasy productions of persons suffering from behavior disorder resemble those of persons not diagnosed ill?

2. Does the restricted environment of a psychiatric hospital affect systematically the fantasy productions of patients?

3. How can the material obtained from the TAT be standardized to furnish adequate norms for use in a clinical situation?

---

[1] The data included in this monograph were collected in part in connection with a Ph.D. dissertation, *An investigation of the fantasy productions of college students and neuropsychiatric patients*, University of Wisconsin, 1949.

[2] The use of specific characteristics of form and content of the subject's productions as pathognomonic indicators for a given clinical syndrome or dynamic factor.

II. PROCEDURE

A. Subjects

The subjects of the present study consisted of 150 male veterans of World War II, divided into six groups of 25 each, which included: Groups C1 and C2, non-hospitalized college students; Group NHPN, non-hospitalized psychoneurotics; Group HPN, hospitalized psychoneurotics;[1] Group HS, hospitalized schizophrenics; Group GNPH, general neuropsychiatric hospital population. The clinical groups contained the following subclassifications (all final diagnoses rendered by a board of at least three psychiatrists):

1. Group NHPN, Non-hospitalized Psychoneurotics:

anxiety	19
mixed	3
anxiety, conversion reaction	1
anxiety with depression	1
obsessive with anxiety	1

2. Group HPN, Hospitalized Psychoneurotics:

anxiety	14
mixed	7
anxiety, conversion reaction	1
anxiety with depression	1
anxiety, sex deviate	1
anxiety, emotional instability	1

3. Group HS, Hospitalized Schizophrenics:

unclassified	14
simple	3
catatonic	3
paranoid	3
hebephrenic	2

4. Group GNPH, General Neuropsychiatric Hospital Population:

schizophrenic, paranoid	5
schizophrenic, catatonic	2
schizophrenic, unclassified	1
schizophrenic, simple	1
incipient schizophrenia	2
manic depressive psychosis	2
psychosis, unclassified	1
psychoneurosis, anxiety	2
psychoneurosis, conversion reaction	1
reactive depression with alcoholism	1
transvestism	1
inadequate personality	1
psychopathic personality	1
drug addiction	1
encephalopathy	1
post epileptic	1
observation	1

The reasons for the breakdown into six groups of 25 each were to both simplify the statistical procedures (i.e., permit the use of frequency data in comparisons without converting to percentages) and to provide experimental control over any group characteristics which might emerge. If significant differences occurred, we could see: (*a*) whether they were more extensive than those between two unsystematically selected normal groups; (*b*) whether hospitalization itself accounted for some differences; (*c*) whether certain group characteristics could emerge even from a heterogeneous population (such as that constituted by Group GNPH). One of the shortcomings of the studies in the use of various tests in clinical diagnosis has been that not only do the diagnostic patterns derived from the examination of test protocols select individuals of the specific clinical group desired, but they also pick out individuals of every other clinical category. In this study we have a check on whether the mere grouping of individuals into clusters of 25 does or does not in itself make for significant group characteristics.

The groups were distributed according to age, education, IQ, and marital status as shown in Table 1. They are comparable with respect to these variables, although Group GNPH is least homogeneous and has more extreme individuals than the other groups.

---

[1] Out-patients seen at Madison VA Mental Hygiene clinic; in-patients seen at Mendota VA Hospital. The writer wishes to thank Dr. R. W. Trent and Dr. A. Bonar for their co-operation in permitting him to see patients at these facilities.

TABLE 1

AGE, EDUCATION, IQ, MARITAL STATUS OF SUBJECTS

Group[a]	N	Age		Education		IQ		Marital Status	
		Range	Mean	Range	Mean	Range	Mean	Married	Divorced
C1	25	20–36	25.9	13–17	14.4			7	
C2	25	20–35	25.8	13–18	14.5			7	
NHPN	25	20–34	25.4	12–16	13.5	103–128	114.2	7	2
HPN	25	19–35	26.1	12–15	13.0	100–126	111.3	15	
HS	25	20–30	24.2	12–17	13.3	100–133	112.9	6	
GNPH	25	20–49	30.6	9–17	12.1	90–128	109.7	9	2

[a] The following notation will be used throughout the Tables and Appendices to identify the various groups:

C1—First group of college students
C2—Second group of college students
NHPN—Group of non-hospitalized psychoneurotics
HPN—Group of hospitalized psychoneurotics
HS—Group of hospitalized schizophrenics
GNPH—Group of general neuropsychiatric hospital patients.

No IQ scores are available for the college students, but it is assumed they are at least of normal intelligence. It has been reported in the literature (46) that, after a normal level of intelligence has been reached, there is no correlation between quantity or quality of response on the TAT and added increments of intelligence. Indeed, one investigator states that stories with satisfactory content for analysis have been obtained with subjects with IQs as low as 80 and consistently satisfactory stories were obtained from subjects with IQ of 90 and above (28). Since the IQ of no subject in the clinical groups is below normal, it is felt that this factor has been adequately controlled.

B. MATERIAL AND ADMINISTRATION

All 20 cards recommended for adult males (Harvard University Press, 3rd Edition) were used. Administration of the test was individual and the only ones present in the testing room were the subject and examiner. The examiner took verbatim notes, as far as possible, of everything that the subject said. Although the authors of the test suggest an interval of one week between the administration of the first and second 10 cards (41), hospital and clinic routine did not permit such a procedure in this study. It was therefore necessary to administer the whole series at one session with a ten-minute break between halves. This method has been recommended by certain investigators (46). Directions to all the subjects were identical and as follows:

> This is a test of imagination. Here are some pictures about which I would like you to make up some stories. You are to tell what happened before, what is happening in the picture, and how it's going to turn out, what the characters are thinking and feeling. Take about four or five minutes for each one but be sure to include all these things. All that we want is a plot for each picture. No elaborate literary masterpiece is necessary. This is the first one, see what you can do with it.

Each subject was prompted when any of the elements of the story as given in the directions was omitted. However, during any single story no subject was prompted more than once for any particular element. After the first ten cards were administered, the subjects were told that the test was half over and that there would be a ten-minute break during which they could go to the lavatory, have a drink of water, etc. After approximately ten minutes the test was resumed with the following instructions:

> The stories which you told in the first half were all very good—just the kind of story I wanted. But the pictures I'm going to show you now are even more difficult to make up stories about. They are more vague and ambiguous, not as clearly defined, and you're really going to have to use your imagination to make up some meaningful stories. So give free rein to your imagination now and make up some good stories about these pictures.

For the sixteenth card, which is blank, there were the special directions as follows:

> "Now look at this blank card. Imagine that you see some kind of picture there. Tell me what it is." After the subject described the picture he was then told, "Now make up a story about it."

C. Technique of Analyzing Stories

After the stories were all collected they were analyzed for the following variables: themes, identification of characters, perceptual distortions, unusual details, level of interpretation, emotional tone of the body of the story and outcome, as follows:

(1) *The thematic analysis* followed a check list of themes empirically derived in the following manner. After 50 protocols had been collected they were perused at least twice. The first time, each story was read through and the action taking place was summarized in a few words. After this was done for all the stories, the author, with the help of another psychologist,[2] examined the summaries and a check list of 98 themes was organized under the two general headings of disequilibrium and equilibrium, indicating the state of tension or adjustment displayed in the story. These two general groups were subdivided into interpersonal, intrapersonal, and impersonal classifications, depending upon the sphere to which the situation was referred. The interpersonal classification was further broken down into sections dealing with parent,[3] partner, peer, and sibling. Each theme appearing in the classification was named and defined and examples of each given, so that the same list could be used in further research. On the basis of additional experience the list was revised and expanded to include 125 themes. The final check list as used in this study, with definitions of each theme, appears in Appendix B. During the examination of the stories, each theme was tallied as it appeared, so that most stories have more than one theme and rarely are there stories with no themes which appear on the check list and which are thus impossible to classify. Only the manifest content was considered—the actual behavior of the characters in the stories as narrated by the subject, regardless of its covert significance. The thematic analysis was done by the experimenter, since it is actually merely a counting procedure with a minimal judgmental function—a theme is either present or not, and the sole criterion is the verbal statement of the subject.

(2) *Identification of characters* included the sex and age of each of the individuals depicted on the cards, the relationship when there were two or more characters, and specific identifying features, including in some cases, nationality, occupation, physical characteristics, etc. In determining perceptual distortions, the standard taken was the description of each picture offered by the authors of the test (41). Any deviations from this standard were noted. Notice was also taken of attention to unusual

---

[2] Dr. Ann Margaret.

[3] Parent substitutes such as older authority figures are included under the heading of "parent." Authority figures of age range similar to hero are classified with "peer" or "partner," whichever is appropriate.

details such as pregnancy of the woman on card 2, lack of tie on man on card 4, etc. Only one judge, the experimenter, was used for identifications of characters, as this too is merely a counting procedure.

(3) *The level of interpretation* refers to whether the responses were narratives, mere descriptions, symbolic, abstract, etc. and also to unusual formal characteristics of the stories such as inclusion of the examiner in the story, no connection between picture and story told, etc. A list of these characteristics and their definitions appears in the latter part of Appendix B. Since the judgmental function is minimal here also only one judge was used.

(4) To rate the *emotional tone of the stories and outcomes,* it was decided to improve upon the three-point scale used in a previous study (19). Consequently a five-point scale was developed by the author and two collaborators (18). The scales which appear in Appendix A were derived in the following manner: 500 TAT stories were read independently by three judges who indicated the rating of each story on a 10 cm. line representing a continuum from happy to sad. Upon completion of the ratings the lines were divided into 10 one cm. intervals and an average Pearson product-moment correlation coefficient of .76 was obtained among the raters. The 10 cm. lines were then divided into five 2 cm. intervals and each interval was assigned a weight as follows, ranging from sad to happy: -2, -1, 0, $+1$, $+2$. Each interval was described on the basis of those stories for which there was complete agreement among the three judges, i.e., all three ratings fell in that particular interval.

Individual scales were constructed for each story for Emotional Tone and Outcome and general scales for the two variables were also devised. The individual scales and the general scale for Outcome were so similar that only the general scale was used in rating the outcome, although the individual scales were used for emotional tone.[4] The inter-interpreter reliability of this scale, based on ratings of three judges for 200 additional stories (10 complete protocols) was rather high. The average Pearson r for emotional tone was .86; and for outcome, .75 (18). It was felt that, since the rating of tone and outcome was a judgmental process, someone other than the investigator should rate the stories as the latter might be biased by a knowledge of the diagnostic classification of the subjects. Therefore independent ratings of the emotional tone and outcome were made by an individual who had no clue as to which were the normal stories and which belonged to members of the various clinical groups.[5]

D. Statistical Method

All the data collected in this study (e.g., judgments of emotional tone, themes, perceptual distortions, etc.) were all of the frequency type and our comparisons, therefore, employ the chi square technique. The formula was one with correction for continuity suggested for use when $p = q = 0.5$ (14).

$$\text{Chi square} = \frac{([\mathrm{m}-\mathrm{s}]-1)^2}{\mathrm{m}+\mathrm{s}}$$

In general each variable was treated in two ways:

(*a*) First the data were analyzed on the basis of total responses, i.e., each group was compared with every other group on the basis of total number of times a given characteristic was present in those groups for all 20 stories. This means that each individual contributes more or fewer responses to the analysis.

---

[4] Individual scales for emotional tone of stories and the general scale for outcome appear in Appendix A.

[5] Mrs. Dorothy T. Hake rated all the stories for emotional tone and outcome.

(*b*) Next the data were analyzed on the basis of individuals, that is the number of persons in each group contributing a given number or more of responses of each type. This provides information concerning whether the factor causing group differences was a greater mean number of responses per individual in the group or just a few extreme individuals. Also if there are no significant differences between groups, the reason might be, not that all individuals in both groups contribute approximately the same number of responses, but that in the one, occasional subjects are responsible for the majority of responses, while in the other all are contributing just a few. In other words we would have some indication of the homogeneity of variance within the groups. Presumably, also, if one group contributed significantly more of certain responses than another, and also contained significantly more individuals who made such responses, we could be more certain that these differences in responses represent real group characteristics. One of the inadequacies of past research on diagnostic testing is that it has been based on responses and not on individuals. This means that any generalization of the conclusions from the groups on whom the work was done to other groups is unjustified since intra-group variability has not been considered.

Chi square tests of homogeneity (20, 33) were also performed in the analysis of individual themes and ratings of tone and outcome. The .05 level of confidence (55) was adopted as the criterion of significance and all differences significant at this level were noted, both in the individual chi square tests and in the tests of homogeneity.

III. RESULTS AND DISCUSSION

A. Emotional Tone of Stories

It is obvious from Table 2 that the preponderance of stories told by all subjects in response to the TAT are sad, although there is a little significant intergroup variation with respect to the proportion of stories in each category of the scale (chi square of homogeneity, 94.9900, with 20 degrees of freedom, significant beyond the .001 level). A comparison of each group of subjects with every other group indicates that the six groups arrange themselves into two distinct populations depending on whether or not they are hospitalized. The hospitalized groups appear to be emotionally duller than the non-hospitalized ones since those in the hospital tell more of the neutral stories and less of the extremely emotional ones than those not in the hospital.

These differences might be a function of the seriousness of the illness since undoubtedly those in the hospital are more seriously ill than the out-patient group. However, the relationship is not always a progressive one, with the non-hospitalized psychoneurotics more like the hospital groups than are the normal subjects, and the psychoneurotics more like the normals than the schizophrenics. There are some instances in which the non-hospitalized psychoneurotics differ significantly from the hospitalized ones and not from the schizophrenics (e.g., total number of sad stories). If degree of illness caused the increasingly neutral stories it would be expected that the schizophrenics would have more neutral stories than the psychoneurotics and that the hospitalized psychoneurotics would be more like the non-hospitalized ones than the schizophrenics. This is not the case and thus it seems that the differences obtained are a function of the hospitalization and not the degree of illness.[1] An analysis of the data according to the number of individuals contributing a given number or more of stories of each category of emotional tone produces the same result. *The hospitalized subjects, regardless of classification, give more stories which are neutral or flat in the amount of feeling invested in them.*

This is the same type of phenomenon that has been noted in individuals exposed to other kinds of restricted, routinized, and deprived environments such as concentration camps, prisons, and orphanages (2, 9, 10, 21, 22, 23, 42,

TABLE 2

Frequency of Stories of Different Emotional Tone, by Group

Tone	Group							Comparison	Chi-square	*P*
	C1	C2	NHPN	HPN	HS	GNPH	Total			
+2	18	15	15	13	13	6	80			
+1	74	60	64	52	55	54	359	−1, 0	186.5721	.001
0	85	97	77	142	159	171	731	0, −2	79.8755	.001
−1	249	254	246	201	205	201	1356	−2, +1	4.9220	.03
−2	74	73	85	79	60	55	426			
Total happy[a]	18.4 92	15 75	15.8 79	13 65	13.6 68	12 60	14.6 439	happy, sad	812.0000	.001
Total sad[a]	64.6 323	65.4 327	66.2 221	56 280	53 265	51.2 256	59.4 1782			

[a] Upper figures are percentages. Lower figures are frequencies.

---

[1] Of course it is understood that the rough diagnostic categories here used are not the best indicators of severity of illness and that to study definitively the effect of hospitalization, duration of stay in the hospital would have to be strictly controlled and systematically varied. However, within the limits of this study it is highly suggestive that our obtained differences are the result of hospitalization and not severity of illness.

51). Reports of observations on such populations all stress the dulling of social adaptation, the apathy, indifference, and aimlessness which such controlled environments engender among their inhabitants. Greenson (23) discusses the cases of men who had been exposed during the war to adverse conditions, often free of actual danger, but with poor food, in bad climate, or with severe boredom and great loneliness. The apathy which develops in such individuals, he believes, is specifically caused by deprivation, which for Greenson refers to "lack or withdrawal of the necessary supplies of love (affection, deference) or food, or both, which the individual requires from his environment." Apathy is the defense against painful perceptions and serves the purpose of avoiding "overwhelming feelings of annihilation."[2]

One possible explanation, thus, of the relatively flattened affect of the stories told by our hospitalized patients is that it is a reflection of the insulation they have built around themselves as a defense against the lack of affective bonds which they experience in this circumscribed environment. It is an effect not limited to just one kind of patient, e.g., schizophrenic, but one which applies to all our hospitalized subjects.

B. OUTCOME OF STORIES

As can be seen from Table 3, the outcomes are more variable for all groups of subjects than the emotional tone of the stories, which seems to be determined largely by the stimulus properties of the pictures themselves. The outcomes appear to be more clearly a function of the individual's creativity.

The most frequent outcome is that described as moderately happy, although the frequency in this category is not significantly greater than that in the neutral category. The fact that story outcomes tend to be happy does not agree with Rautman and Brower's findings with children (45), although it does with Coleman's investigations with children (15), and with a former study done with adults (19). Rautman and Brower, however, worked with only 10 cards, three of which

TABLE 3

FREQUENCY OF STORIES OF DIFFERENT OUTCOMES, BY GROUPS

Outcome	Group							Comparison	Chi-square	p
	C1	C2	NHPN	HPN	HS	GNPH	Total			
+2	45	27	33	16	17	18	156			
+1	174	158	143	148	143	155	921			
0	120	147	108	135	172	162	844			
−1	81	71	58	35	49	37	331	0, ?	180.5954	.001
−2	56	67	76	48	40	39	326	+2, −2	59.2552	.001
?	24	29	69	105	71	76	374			
Total happy[a]	42.4 219	37.7 185	35.2 176	32.8 164	32.0 160	35.5 173	35.9 1077			
Total sad[a]	27.4 137	27.6 138	26.8 134	16.6 83	17.8 89	15.2 76	21.9 657	happy, sad	101.2462	.001

[a] Upper figures are percentages, lower figures are frequencies.

---

[2] It is interesting that the prewar histories in his group of patients showed that men from broken homes and orphan asylums formed a relatively high percentage of those who developed apathy.

are not in this series, and administered the test in group fashion. Thus their results are not strictly comparable to those of the present study. However, they do mention the fact that happy endings were more common among teachers who took the test than with the children, but give no indication of the extent of the difference. There might thus be a difference between children and adults on this variable.

There is also much more intergroup variability in relation to outcome than was found in a consideration of emotional tone (chi square of homogeneity = 184.61 with 25 degrees of freedom, significant far beyond the .001 level of confidence). The differences between the hospitalized and non-hospitalized groups are not as clearly defined as they were in the previous analysis, although the hospitalized groups give more neutral and less extreme outcomes than do the non-hospitalized groups. In addition, each non-hospitalized group gives significantly more sad outcomes than each hospital group. When the data are analyzed by the number of individuals in each group contributing a certain number or more of a given outcome, the pattern is the same—fewer individuals in the hospitalized groups give sad outcomes and fewer also give extreme outcomes. Significantly more non-hospitalized psychoneurotics give a large number of moderately sad outcomes than do hospitalized psychoneurotics and, as in the analysis of emotional tone, it would thus seem that hospitalization itself is a factor in making the outcomes more bland.[3]

Both normal groups give significantly less questionable outcomes than each of the clinical groups. Of all the clinical groups, the hospitalized psychoneurotics give the most such conclusions, significantly more than both the non-hospitalized psychoneurotics and the schizophrenics. Essentially the same picture is obtained when the data are analyzed by individuals. No normal subject gives more than four such outcomes, whereas in the clinical groups there are individuals contributing as many as 12. This excess of questionable outcomes among all the clinical groups is perhaps a reflection of the indecisiveness and uncertainty of these patients who are not as forthright and sure of themselves as are normal individuals. This is carried to an extreme by the hospitalized psychoneurotics.

C. SHIFT IN TONE FROM BODY OF STORY TO OUTCOME

The fact that many individuals turn unhappy stories into happy conclusions has been noted before (15, 19) and the frequency of this phenomenon for all our groups of subjects appears in Table 4.

It has been stated that the shift to happier conclusions is characteristic of schizophrenic individuals (26). Our data would demonstrate it to be more representative of normal persons than it is of schizophrenics. Another writer has characterized it as a "refusal to accept consequences which (in children) may lay the foundations for later neurotic behavior" (45, p. 200). Actually our normal subjects also manifest this characteristic more frequently than do the psychoneurotics, significantly so in the case of the outpatient group. That individuals often tack illogically happy endings on to very sad stories is not surprising since, by the mere instructions to make up a story, we are inviting the subject to indulge in fantasy. And one of the characteristics of fantasy is that it disregards logic and

---

[3] See note on page 14.

TABLE 4

FREQUENCY OF SHIFT IN TOTAL NUMBER OF STORIES

Direction and amount of shift	Group							Significant comparisons	Differences (chi-square)	P
	Total	C1	C2	NHPN	HPN	HS	GNPH			
+4	9	3	2	1	3					
+3	127	31	23	22	17	17	17			
+2	588	111	103	100	98	81	95	C1-HS	4.3802	.04
+1	409	69	78	47	60	77	78	NHPN-C2	7.2000	.01
								NHPN-GNPH	7.2000	.01
								NHPN-HS	6.7280	.015
All stories shifting positively[a]	37.1	42.8	41.2	34	35.6	35	38			
	1133	214	206	170	178	175	190	C1-NHPN	4.8150	.03
Shifting 2 or more places positively[a]	24.1	29	25.6	24.6	23.6	19.6	22.4			
	724	145	128	123	118	98	112	C1-GNPH	3.9874	.05
								C1-HS	8.7078	.005
Shifting 3 or more places positively[a]	4.5	6.8	5	4.6	4	3.4	3.4			
	136	34	25	23	20	17	17	C1-HS	5.0196	.025
								C1-GNPH	5.0196	.025
0	1578	237	245	269	270	288	269	C1-HS	4.9543	.03
−1	198	39	38	35	35	26	25			
−2	29	7	8	5	3	3	3			
−3	14	3	2	8	1			NHPN-HS	4.0000	.05
All stories shifting negatively[a]	8.0	9.8	9.6	9.6	7.8	5.8	5.6			
	241	49	48	48	39	29	28	C1-GNPH	5.1948	.025
								C2-GNPH	4.7500	.03
								NHPN-GNPH	4.7500	.03
								C1-HS	4.6282	.035
								C2-HS	4.2078	.04
								NHPN-HS	4.2078	.04
								hosp. vs. non hosp.	9.5210	.005
All stories shifting 2 or more places negatively[a]	1.4	2.0	2.0	2.6	.1	.1	.1			
	43	10	10	13	4	3	3	NHPN-HS	5.0625	.025
								NHPN-GNPH	5.0635	.025

[a] Figures appearing in top row are percentages. Figures in bottom row represent frequencies.

reason (39). Furthermore, the individual is following a cultural pattern illustrated by our movies which, producers seem to feel, must always have happy endings in order to be successful financially and which thus reflect the demand of the moviegoers.

D. Perceptual Distortions

1. *Misidentification of Sex*

Instances in which the subject misidentifies the sex of characters in the pictures are frequent and there is no significant difference among the clinical and normal groups in the total number of such misidentifications (Table 5). Nor is there a significant difference in the number of individuals in each group making them. There are certain pictures in which sexual misidentification is frequent; others where it is completely absent. Some errors in sex identification are made by individuals in each group; others, only by individuals in the clinical groups; and one, only by a normal individual.

Many clinicians interpret the fact that a subject misidentifies the sex of a character in the picture as an indication of his own sex confusion. Numerous writers (7, 26, 44, 46, 52) have specifically indicated that sexual misidentification of the character on Card 3BM is an indication of a high feminine component in the personality of the male subject. However, more than half the subjects are confused over the sex of this character and this corroborates an earlier investigation (16). It is difficult to see how this can legitimately be called a projection of the subject's feminine identification when fewer individuals identify it correctly than err in identification. This too would seem to be more a function of the stimulus properties of the card. There may be some justification in such a statement about the other pictures, where misidentification is rare, and where it can be assumed there is actual distortion on the part of the subject when he misidentifies a character. It should be noted, however, that among our subjects are 7 homosexuals, ranging from one diagnosed as latent homosexual to one who has been a transvestite for a number of years, and in no one of these cases is there a single instance of sexual misidentification.

TABLE 5

Frequency of Instances of Sex Confusion[a]

Card	Group						
	C1	C2	NHPN	HPN	HS	GNPH	Total
3BM	15	12	19	12	12	15	85
5		1					1
6BM					1		1
8BM	1	1	1	1	1	2	7
10	1			4	1	2	8
12BM	3	2	4	3	2	6	20
14					1	1	2
15	1		2		1	2	6
20	6	4	4	3	5	5	27
Total	27	20	30	23	24	33	157
Percentage	5.4	4	6	4.6	4.8	6.6	5.0

[a] This includes outright misidentification and instances in which subject is unsure whether character is male or female. For exact nature of these confusions and all perceptual distortions, refer to Appendix C.

2. *Other Perceptual Distortions*

Although the most frequent perceptual distortions are the sexual misidentifications, Table 6 shows that other distortions are also not infrequent and occur in each group. By "other perceptual distortions" are meant any departures from the standard description of the picture as described in the test manual (41). These include such things as identifying the violin in card 1 as a typewriter, the gun in card 3 as a pipe, the woman in card 4 as a negress.[4] Again there are some cards for which distortions are frequent in all groups and other cards in which perceptual distortions are made only by members of the clinical groups. The general hospital group produces significantly more distortions than all other groups. But there are no significant differences among any of the groups as to number of individuals producing these distortions. From 70% (Psychoneurotic in-patients) to 100% (General hospital group) of the individuals in each group produce at least one distortion and the schizophrenics produce significantly no more than do the normals or psychoneurotics.

A common characteristic ascribed to TAT productions of schizophrenic patients has been frequent distortion of objective characteristics of the pictures (4, 44, 46). However, these data show that perceptual distortions are no more representative of our group of schizophrenics than they are of any of our other subjects, normal or ill. This is evidence of Cameron's "principle of continuity" (13). Distortions of reality are common among normal individuals as well as among persons with behavior disorders. The only difference, important as it may be, is that normal people are better able to check up on their faulty perceptions through appropriate and practiced tech-

TABLE 6

FREQUENCY OF OTHER PERCEPTUAL DISTORTIONS

Card	Groups							Significant Intergroup Differences		
	C1	C2	NHPN	HPN	HS	GNPH	Total	Comparisons	Chi-square	*P*
1			1	3	1	1	6			
3BM	7	13	5	5	1	14	45			
4	3	6	2	1	6	3	21			
6BM					3	2	5			
7BM			1		1		2			
8BM	12	8	9	9	12	18	68			
9BM	3	6	2	5	4	3	23			
12BM	3	1	5	4	4	6	23			
13MF	1		3		5	6	15			
14				2	2	5	9			
15	3				1	5	9			
17BM	1				2	3	6			
18BM	6		3	2	5	5	21			
20					1	1	2			
Total	39	34	31	31	48	72	255	C1-GNPH	7.2900	.01
Percentage	7.8	6.8	6.2	6.2	9.8	14.4	8.5	C2-GNPH	10.0104	.005
								HPN-CNPH	10.7895	.005
								NHPN-GNPH	10.0104	.005
								HS-GNPH	4.8090	.03[a]

[a] No significant differences when analyzed by individuals.

---

[4] For instances of specific distortions see Appendix C.

niques of social validation. The critical point in the development of delusions, e.g., is, in addition to the need and the "reaction sensitivity," the failure of correction.

It is difficult to see, on the basis of our data, how perceptual inaccuracies can be used as a diagnostic "cue" for schizophrenia. Of course, the schizophrenics in our sample are all young, just recently diagnosed, and with comparatively little deterioration. However, these are the cases which are the diagnostic problems. The TAT is not necessary in order to ascertain that a deteriorated schizophrenic has bizarre ideas or distorts reality. But when the TAT is used as a diagnostic instrument with suspected cases of schizophrenia it is unjustifiable to employ presence or absence of perceptual distortions as a "sign."

E. Themes

In the approximately 3,000 stories which were analyzed in this study there were 6,210 discernible themes which could be classified according to our check list. These were contributed by the various groups as indicated in Table 7. All the groups, except (GNPH), are about equally productive, with no significant difference in the number of themes produced. The last group contributes significantly less than each of the other groups, except (HPN), and the difference, even in this comparison (chi square = 3.7809), approaches significance at our criterion. It is obvious that lack of productivity is not a characteristic of any of our more homogeneous groups (indeed, in average number of responses per individual there is little difference among any of the groups). Schizophrenics as a class do not show an unusual[5] amount of blocking as has been reported (4, 8).

TABLE 7

Number of Themes Contributed by Subjects in Each Group

Group	No. of Themes	Average No. per Individual
C1	1,073	43
C2	1,088	44
NHPN	1,067	43
HPN	1,017	41
HS	1,034	41
GNPH	961	38

As for the themes themselves, the 60 most popular ones, including all those appearing at least 30 times (or in approximately 1% of the stories), are listed in Table 8. The inspection of this table makes it apparent that themes of violence, hostility, death, restriction, guilt, and frustration are common to subjects of all types in response to the TAT and are not peculiar to any particular clinical group as has been reported in the literature (5, 8, 25, 30, 31, 32, 35, 44, 47, 50). Themes for which these subjects do not comprise a homogeneous population appear in Table 9. There are 11 such individual themes in our classification scheme of 125 themes, which is more than could be expected by chance at the .05 confidence level. However, when individual comparisons between groups are considered, only 72 significant differences, listed in Table 10, occur out of a total of 1,875 comparisons (125 themes × 15 comparisons per theme). This is less than would be expected by chance 5% of the time. It should be noted that significance could not be calculated by the chi square procedure in 321 comparisons because there are 45 themes for which one or more groups have a zero frequency. But

---

[5] By unusual is meant a difference in frequency which is statistically significant at the .05 level of confidence.

TABLE 8

MOST FREQUENT THEMES CONTRIBUTED BY 150 SUBJECTS IN 3,000 *TAT* STORIES[a]

	Theme[b]		Group						
	No.	Name	C1	C2	NHPN	HPN	HS	GNPH	Total
1	IA1a	Pressure from parents	51	65	50	31	58	34	289
2	IC5	Aggression from impersonal source	43	49	47	39	45	44	267
3	IB1	Aspiration	46	54	45	46	41	32	264
4	IB3	Curiosity	34	41	32	20	38	42	207
5	IA1b	Succorance from parents	27	29	32	33	30	35	186
6	IB4	Behavior disorder	33	31	33	26	28	17	168
7	IB19	Vacillation	27	31	26	25	28	29	166
8	IA2j	Death or illness of partner	18	23	29	39	33	20	162
9	IB11	Occupational dissatisfaction	41	19	27	22	28	20	157
10	IB12	Physical illness or death of central character	20	22	23	31	24	31	151
11	IC1	Economic pressure	29	26	21	27	27	17	147
12	IC6	War	29	31	28	24	12	22	146
13	IIB6	Occupational satisfaction	16	20	27	24	32	27	146
14	IB7	Guilt—remorse	28	24	18	34	18	21	143
15	IA3b	Succorance from peer	24	19	22	14	23	32	134
16	IB8	Drunkenness	21	26	22	27	19	17	132
17	IA2a	Pressure from partner	23	20	25	19	17	22	126
18	IB16	Religion	15	16	18	25	30	22	126
19	IC4	Aggression towards environment	16	22	22	23	22	18	123
20	IB2	Inadequacy	17	23	17	23	15	13	108
21	IA2h	Illicit sex	25	29	14	15	14	7	104
22	IA1i	Death or illness of parent	22	9	22	15	14	13	95
23	IIA2c	Contentment of partner	16	12	13	17	19	15	92
24	IA3j	Belongingness	22	23	16	11	9	7	88
25	IC2	Legal restriction	6	11	10	29	18	14	88
26	IA1f	Departure from parent	16	12	20	7	18	11	84
27	IA3e	Aggression to peer	11	14	15	15	20	8	83
28	IC7	Escape from perilous environment	14	11	12	16	12	10	75
29	IB13	Retribution	10	9	11	16	15	10	71
30	IB5	Suicide	4	11	13	15	18	9	70
31	IB17	Loneliness	15	9	9	14	8	12	67
32	IB9	Fear	13	14	11	7	13	8	66
33	IIB2	Intrapersonal tranquility	15	13	13	7	6	9	63
34	IA3m	Hypnotism	14	7	9	12	6	13	61
35	IA1j	Death or illness of child	18	13	9	3	10	7	60
36	IA3d	Aggression from peer	6	9	17	9	12	6	59
37	IB21	Exhaustion	7	6	9	8	11	17	58
38	IA3i	Death or illness of peer	11	8	10	8	10	10	57
39	IB10	Rumination	15	18	6	5	6	7	57
40	IA2f	Departure from partner	8	15	8	7	8	9	55
41	IIA1f	Contentment with parent	10	8	8	1	9	11	53
42	IA3a	Pressure from peer	11	12	6	9	5	9	52
43	IC3	Generalized restriction	17	14	7	5	4	4	51
44	IIB3	Reminiscence, happy	10	7	10	12	4	8	51
45	IA2B	Succorance from partner	6	7	6	11	8	7	45
46	IA2e	Aggression towards partner	4	8	9	12	8	2	43
47	IIB1	Self esteem	9	10	8	6	3	7	43
48	IIB7	Ordinary activity intra-personal	5	7	4	5	7	14	42
49	IA1m	Filial obligation	8	8	8	3	6	8	41
50	IA2n	Partner competition	3	4	8	7	7	9	38
51	IIC1	Favorable environment	9	8	5	8	7	1	38
52	IIA3e	Exhibition	8	7	6	3	4	8	36
53	IA1p	News of marriage to parent	5	11	2	6	5	6	35
54	IB14	Reminiscence, sad	6	6	6	3	8	5	34
55	IA1g	Parental concern	5	3	5	7	7	6	33
56	IA2c	Nurturance to partner	7	6	8	4	3	5	33
57	IA31	Competition with peers	10	6	4	2	5	6	33
58	IB6	Moral struggle	2	6	9	4	4	5	30
59	IA2m	Jealousy of partner	5	10	3	6	3	3	30
60	IA1k	Disappointment to parent	7	5	2	7	4	5	30

---

[a] Included are all themes appearing in 1% (30) or more of the 3,000 stories. Themes are listed in order of decreasing frequency.

[b] For definitions of specific themes see Appendix B.

TABLE 9

THEMES FOR WHICH ALL GROUPS OF SUBJECTS DO NOT MAKE UP A HOMOGENEOUS POPULATION AT .05 CONFIDENCE LEVEL

Theme		Chi-square of homogeneity	*P* (5 degrees of freedom)
No.	Name		
IA1a	Pressure from Parent	19.2290	.01
IA1j	Death or Illness of Child	13.2000	.05
IA1c	Nurturance to Parent	16.2500	.01
	Total Themes Parental Disequilibrium	14.9178	.02
IA2h	Illicit Sex	21.5294	.001
IA2j	Death or Illness of Partner	12.2221	.05
	Total Themes Partner Disequilibrium	11.4887	.05
IA3j	Belongingness	15.3335	.01
	Total Themes Interpersonal Disequilibrium	17.0565	.01
IB10	Rumination	27.1578	.001
IB11	Occupational Dissatisfaction	12.7306	.05
IC2	Legal Restriction	21.8668	.001
	Total Themes of Disequilibrium	21.5186	.001
IC3	Generalized Restriction of Environment	20.3293	.01
IIA1d	Reunion with Parents	12.0549	.05
	Total No. of Themes	16.3409	.01

even on the basis of the remaining 1,554 comparisons, there are fewer significant differences than could be expected by chance. Thus it would seem that, at least on the basis here proposed, content analysis cannot differentiate diagnostic classifications.

In order to say that any given feature characterizes a certain clinical group, this feature *must distinguish that group both from normal individuals and from other clinical groups.* Not one of our classes of subjects differs significantly from *all* other classes in any given theme. Furthermore, no class differs from *both* normal groups in more than five themes, and there are no themes for which the non-hospitalized psychoneurotics differ from both normal groups. In fact there are fewer significant differences between one normal group (C1) and one clinical group (NHPN) than there are between the two normal groups (C1) and (C2). By chance there should be more than six such differences in 125 comparisons. There are only five pairs of groups that have more than six significant differences between them and thus differ more than could be expected by chance in regard to themes. These are (C1) and (HPN) with 10 significant differences, (C1), and (GNPH) with seven significant differences, (C1) and (HS) with eight significant differences, (C2) and (HPN) with seven significant differences, and (C2) and (GNPH) with eight significant differences.

As was mentioned above, there are 45 themes in which the frequency for at least one group is zero and therefore 321 comparisons could not be made. The maximum number of such instances for any particular comparison, however, is 28, and when the total number of comparisons between any two groups is reduced by the number of times that the chi square could not be calculated, the number of differences more frequent than could be obtained by chance between any two classes is not greatly affected. The comparison between another pair of groups, (HPN) and (GNPH) can now be added to the above list. It would seem therefore, that we could say definitely no more than that hospitalized psychoneu-

TABLE 10

Differences in Themes and Interpretation Levels Significant at or Beyond .05 Confidence Level

Theme		Group frequency		Group frequency		Chi-square	P
No.	Name						
IA1a	Pressure from Parent	C2	65	HPN	31	11.3437	.001[b]
		C2	65	GNPH	34	9.0909	.005
		HS	58	HPN	31	7.5955	.01
		HS	58	GNPH	34	5.7500	.02
		C1	51	HPN	31	4.4024	.035
		NHPN	50	HPN	31	4.0000	.05
IA1c	Nurturance to Parent	C1	9[a]	HPN	1	4.9000	.03
		C1	9[a]	HS	1	4.9000	.03
		C2	8[a]	HPN	1	4.0000	.05
		C2	8[a]	HS	1	4.0000	.05
IA1f	Departure from Parent	NHPN	20	HPN	7	5.3333	.025
IA1i	Death or Illness of Parent	C1	22	C2	9	4.6451	.035
		NHPN	22	C2	9	4.6451	.035
IA1j	Death or Illness of Child	C1	18	HPN	3	9.3333	.005[b]
		C2	13	HPN	3	5.0625	.025
		C1	18	GNPH	7	4.0000	.05
IA1p	Marriage of Child	C2	11[a]	NHPN	2	4.9231	.03
IA2e	Aggression to Partner	HPN	12	GNPH	2	5.7857	.02
IA2h	Illicit Sex	C2	29	GNPH	7	12.2500	.001[b]
		C1	25	GNPH	7	9.0312	.005[b]
		C2	29	NHPN	14	4.5581	.035[b]
		C2	29	HS	14	4.5581	.035
		C2	29	HPN	15	3.8409	.05
IA2j	Death or Illness of Partner	HPN	39	C1	18	7.0175	.01[b]
		HPN	39	GNPH	20	5.4916	.02
		HS	33	C1	18	3.8431	.05[b]
IA3b	Succorance from Peer	GNPH	32	HPN	14	6.2826	.015
IA3e	Aggression to Peer	HS	20	GNPH	8	4.3213	.04
IA3d	Aggression from Peer	NHPN	17[a]	C1	6	4.3478	.04
		NHPN	17[a]	GNPH	6	4.3478	.04
IA3j	Belongingness	C2	23	GNPH	7	7.5000	.01
		C1	22	GNPH	7	6.7586	.005
		C2	23	HS	9	5.2812	.025
		C1	22	HS	9	4.6451	.035[b]
IA3l	Competition	C1	10[a]	HPN	2	4.0833	.045
IB1	Aspiration	C2	54	GNPH	32	5.1279	.025[b]
IB3	Curiosity	GNPH	42	HPN	20	7.1129	.01
		C2	41	HPN	20	6.5574	.015
		HS	38	HPN	20	4.9828	.03

[a] No individual contributes more than one such response.

[b] Also a difference in number of individuals in each group.

(*Continued on next page*)

Theme		Group frequency		Group frequency		Chi-square	*P*
No.	Name						
IB4	Behavior Disorder	C1	33	GNPH	17	4.5000	.035
		NHPN	33	GNPH	17	4.5000	.035
IB5	Suicide	HS	18	C1	4	7.6818	.01
		HPN	15	C1	4	5.2632	.025[b]
IB7	Guilt—Remorse	HPN	34	NHPN	18	4.3269	.04
		HPN	34	HS	18	4.3269	.04
IB10	Rumination	C2	18	HPN	5	6.2609	.015
		C2	18	NHPN	6	5.0417	.025
		C2	18	HS	6	5.0417	.025
		C1	15	HPN	5	4.0500	.045
		C2	18	GNPH	7	4.0000	.05
IB11	Occupational Dissatisfaction	C1	41	C2	19	7.3500	.01
		C1	41	GNPH	20	6.5574	.015
		C1	41	HPN	22	5.1428	.025
IB21	Exhaustion	GNPH	17[a]	C2	6	4.3478	.04
IC2	Legal Restriction	HPN	29	C1	6	13.8286	.001[b]
		HPN	29	NHPN	10	8.3077	.005
		HPN	29	C2	11	7.2250	.01
		HPN	29	GNPH	14	5.9394	.015
		HS	18	C1	6	5.0416	.025
IC3	Generalized Restriction	C1	17	GNPH	4	6,8570	.01
		C1	17	HS	4	6.8570	.01[b]
		C1	17	HPN	5	5.5000	.02
		C2	14	GNPH	4	4.5000	.035
		C2	14	HS	4	4.5000	.035
IC6	War	C2	31	HS	12	7.5348	.01
		C1	29	HS	12	6.2439	.015
		NHPN	28	HS	12	5.6250	.02
IIB4	Retirement	HS	32	C1	16	4.6875	.035
IIB7	Ordinary Activity	GNPH	14	NHPN	4	4.5000	.035
IIC1	Favorable Environment	C1	9	GNPH	1	4.9000	.03
		C2	8	GNPH	1	4.0000	.05
		HPN	8	GNPH	1	4.0000	.05
LEVEL OF INTERPRETATION							
III A	Symbolic	C1	22	NHPN	6	8.0357	.005
		C2	22	NHPN	6	8.0357	.005
		C1	22	HPN	8	5.6333	.02
		C2	22	HPN	8	5.6333	.02
		HS	18	NHPN	6	5.0417	.025
III C	Description	GNPH	21	C2	3	12.0417	.001
		GNPH	21	C1	4	10.2400	.005
		GNPH	21	NHPN	5	8.6538	.005
		HPN	15	C2	3	6.7222	.01[b]
		HS	14	C2	3	5.8823	.02
		HPN	15	C1	4	5.2631	.025[b]
		HS	14	C1	4	4.5000	.03
		HPN	15	NHPN	5	5.0500	.025
III D	Unreal	C1	19	GNPH	6	5.7600	.02

(*Continued on next page*)

Theme		Group frequency		Group frequency		Chi-square	P
No.	Name						
III G	Autobiographical	HPN	25	C1	12	3.8919	.05
III I	Alternate Themes	HPN	45	C1	11	19.4464	.001[b]
		NHPN	35	C1	11	11.5000	.001[b]
		GNPH	31	C1	11	8.5952	.005
		HS	30	C1	11	5.5609	.02[b]
		HPN	45	C2	23	6.4853	.015
III J	Comments	GNPH	20	NHPN	4	9.3750	.005
		GNPH	20	C2	6	6.5000	.015
		HS	16	NHPN	4	6.0500	.015
		GNPH	20	HPN	8	4.3214	.04
III L	Rejection	HS	12	C2	1	7.6923	.01[b]
		GNPH	9	C2	1	4.9000	.03
		HS	12	NHPN	2	5.7857	.02
		HPN	22	C2	1	17.3913	.001[b]
		HPN	22	NHPN	2	15.0417	.001
III M	Peculiar Verbalizations	HS	13	HPN[c]	1	8.6429	.005
III N	Confused	HS	18	C1	1	13.4737	.001
		HS	18	NHPN	1	13.4737	.001[d]
		GNPH	11	C1	1	6.7500	.01
		GNPH	11	NHPN	1	6.7500	.01[b]

[c] All other groups except GNPH have zero frequency. GNPH has frequency four which is not significantly different from 13.

[d] Groups C2 and HPN have zero frequency.

rotics tell certain kinds of themes which are distinctive, since this is the only group that differs from both normal groups and from a miscellaneous neuropsychiatric group more frequently than could be expected by chance.

When the data are analyzed by individuals, the latter statement must be even more circumscribed because on this basis significant differences are still less frequent. Now there are significant differences merely between each of the normal groups and each of the clinical groups; but the number of these differences ranges from only one to three—less than could be expected by chance. Thus it can justifiably be said only that this particular group of hospitalized psychoneurotics differs significantly more frequently than could be expected by chance from these particular normal groups and this particular general neuropsychiatric hospital group.

It is possible to rearrange the subjects so that there are larger classes and thus the comparisons can be based on more individuals as in Table 11. When all psychoneurotics, Groups (NHPN) and (HPN), are compared with all normals, Groups (C1) and (C2), more significant differences appear. There are now 11 such differences, almost twice what could be expected by chance. And when, as in Table 12, all hospitalized groups, (HPN), (HS), and (GNPH) are compared with all non-hospitalized groups (C1), (C2), and (NHPN), slightly more differences emerge. It may be that our original six classes are not large enough to permit significant group differences to emerge.

Some of the differences which are obtained when the subjects are regrouped

TABLE 11

DIFFERENCES IN THEMES SIGNIFICANT AT OR BEYOND .05 CONFIDENCE LEVEL BETWEEN ALL NORMAL AND ALL PSYCHONEUROTIC SUBJECTS

Theme		Normal	Psychoneurotic	Chi-Square	*P*
No.	Name				
IA1a	Parental Pressure	116	81	5.8680	.02
IA1c	Nurturance to Parent	17	4	6.8751	.01
IA1j	Death or Illness of Child	31	12	7.5344	.01
	Total Parental Disequilibrium	356	290	6.5402	.02
IA2h	Illicit Sex	54	29	7.5300	.01
IA2j	Death or Illness of Partner	41	68	6.2018	.02
IA3j	Belongingness	45	27	4.0139	.05
IB10	Rumination	33	11	10.0227	.01
IB11	Occupational Dissatisfaction	60	39	4.0404	.05
IC2	Legal Restriction	17	39	7.8751	.01
IC3	Generalized Restriction of Environment	31	12	7.5349	.01
IIA1d	Reunion with Parents	2	10	4.0833	.05
Level					
III A	Symbolic	44	14	14.5000	.001
III C	Descriptive	7	20	5.3333	.05
III E	Fairy Tale	8	1	4.0000	.05
III I	Alternate Themes	34	80	17.7631	.001
III L	Rejection	1	24	19.3600	.001

TABLE 12

DIFFERENCES IN THEMES SIGNIFICANT AT OR BEYOND .05 CONFIDENCE LEVEL BETWEEN HOSPITALIZED AND NON-HOSPITALIZED SUBJECTS

Theme		Non-Hospitalized	Hospitalized	Chi-Square	*P*
No.	Name				
Ia1a	Parental Pressure	166	123	6.1038	.02
IA1c	Nurturance to Parent	20	5	7.8400	.01
IA1f	Departure from Parent	48	26	5.9595	.02
IA1j	Death or Illness of Child	40	20	6.0167	.02
	Total Parental Disequilibrium	521	423	9.9671	.01
IA2h	Illicit Sex	68	36	9.2404	.01
IA3j	Belongingness	61	28	11.5102	.001
	Total Peer Disequilibrium	348	289	5.2810	.05
	Total Interpersonal Disequilibrium	1290	1127	10.9913	.001
IB2	Inadequacy	57	88	6.2069	.02
IB10	Rumination	39	18	7.0175	.01
IB16	Religion	49	77	5.7857	.02
IC2	Legal Restriction	27	61	12.3750	.001
IC3	Generalized Restriction of Environment	38	13	11.2941	.001
IC6	War	88	58	5.7603	.02
	Total Disequilibrium	2864	2609	11.8810	.001
IIB2	Tranquility	41	22	5.1428	.05
	Total Themes	3228	2982	9.7449	.01
Level					
III A	Symbolic	50	28	4.7619	.05
III C	Descriptive	12	50	12.2208	.001
III G	Autobiographical	41	68	6.2018	.02
III I	Alternate Themes	69	106	6.6057	.02
III J	Comments	20	44	8.2656	.02
III L	Rejections	3	43	33.0652	.001

into these larger classifications are interesting. The normal subjects give significantly more themes of parental disequilibrium than do the psychoneurotics, and within this category, more of both pressure from parents and nurturance to parents. These normals apparently project their parental difficulties into their fantasy more frequently than do these psychoneurotics. However, theirs is perhaps a more mature level of fantasy because a greater number of themes of nurturance to the parent may be an expression of independence, even while more themes of parental pressure reflects a chafing at parental restrictions.

That the normal subjects contribute more themes concerning illicit sex may be due to fewer inhibitions in giving verbal expression to these thoughts—or, if one accepts the concept of need projection, perhaps they have a greater impulse to express them verbally since they are finding less expression actually. The fact that they do not contribute as many themes of death of partner possibly indicates less hostility towards the opposite sex, which might very well be a function of the fewer inhibitions just noted. The normal subjects offer more themes of belongingness, perhaps because, as college students, they are more concerned with problems of acceptance and fellowship; and the traditional preoccupation of college students with world problems and man's insignificance may be reflected in the preponderance of themes of rumination. Bennett (8) also found sexual topics and "disillusionment" (similar to our "rumination") more popular with normals.

Some of these same differences appear when hospitalized and nonhospitalized subjects are compared; and new ones of interest are also revealed. Those in the hospital give significantly fewer themes of interpersonal disequilibrium, including fewer themes of disturbance in the parental situation and fewer themes of illicit sex and belongingness. In general, they give fewer themes of disequilibrium, perhaps as a result of their simplified, protected environment which removes them from many everyday struggles. This is further illustrated by the fact that they have fewer themes of generalized restriction of the environment, yet more of legal restriction, the latter probably a reflection of their own position in the hospital. They also seem to be less concerned with war than the other veterans. The hospitalized subjects offer significantly more themes of inadequacy, perhaps because of their own inferiority feelings which are intensified by the hospital experience. More religious themes might be expected of more seriously ill neuropsychiatric patients, while fewer themes of rumination might indicate a lessened concern with other philosophical problems.

Although there is no indication of the significance, because chi square tests could not be performed, it is important to note those instances where frequency of theme in one group was zero. This might indicate certain characteristics which are exclusively clinical or exclusively normal, at least for this sample. We find that no subject in any of the non-hospitalized groups offers a theme of homosexuality, while at least one such response is forthcoming from each of the hospitalized groups and five from the schizophrenic classification. However, in the latter group there are only two individuals who contribute such themes, since one contributes four of the five themes. Nor does a single theme of incest with siblings appear among the normal subjects, although there are two such schizophrenic responses and one in the non-hospitalized psychoneurotic group. However, one theme of incest with parents is contributed by a normal, and one by a hospitalized psychoneurotic, but none by a schizophrenic or any other clinical subject. It cannot be said then that tabooed sexual themes (incest, rape, homosexuality, etc.) are exclusively schizophrenic as has been stated in the literature (44).

It is interesting that although each group has at least one response of disregard for peer, there is no such response in the schizophrenic group. The same is true of the theme, self pity. In general there is only an occasional theme which is found exclusively in schizophrenic or exclusively in psychoneurotic or exclusively in normal groups. For in almost every classification to which one group contributes, some members of at least one other group also appear.

Although we have hypothesized about the significance of some of the intergroup differences, *the outstanding finding of this study seems to be the similarity in fantasy content among our various groups of subjects.* This is not at variance with what has been reported by some investigators using other projective techniques. More than 25 years ago Gardner Murphy (37) reported that he could not establish any clear-cut differences among the responses of dementia praecox, manic depressive, and paretic patients to the Kent-Rosanoff word list on the basis of preponderance of one type of association. A further study (38) of dementia praecox and manic depressive patients and normal subjects, classifying their responses according to the logical relationship between S and R, also failed to differentiate among these groups. McDowell (34) also found no difference between stuttering and non-stuttering children on the basis of the word association test.

Anastasi and Foley (3) found extensive overlapping in the type of subject matter portrayed by abnormal and normal groups in an investigation of their spontaneous drawings. Their findings are very similar to ours. They say, "Most of the subject matter categories do not differentiate significantly between normal and abnormal drawings. Although certain types of drawings appear to be more clearly diagnostic of abnormality, such drawings are made by a small minority of institutionalized patients. The majority of abnormal patients, although showing clearly psychotic behavior in other respects produced drawings which were indistinguishable from the normal in subject matter" (3, p. 174).

Recent studies have been carried out on populations similar to ours using the Draw-A-Man technique. Royal (48) found no statistically significant differences between neurotic and normal veterans on a check list of 28 drawing characteristics, although there was a tendency towards significance in eight of them. Albee and Hamlin (1) found that 15 clinical psychologists could not distinguish the drawings of psychoneurotic and schizophrenic out-patient veterans from each other.

Using the verbal summator technique, Grings (24) found no discrimination in projective content among 24 schizophrenics, 15 manic depressive, depressed patients and 18 psychoneurotics.

The fact that there are no differences in the thematic content of individuals of the various nosological groups has implications for the theory of behavior pathology. It is further evidence for the hypothesis of continuity between normal and abnormal. According to J. F. Brown, "abnormal psychological phenomena are simply exaggerations (i.e., overdevelopments or underdevelopments) or disguised (i.e., perverted) developments of normal psychological phenomena" (11, p. 7). White (53) has pointed out how maladjustments arise out of each stage or process of normal growth, and Cameron (13) contends that normals, neurotics, and psychotics differ from one another only in degree and that any behavior found in psychiatric patients is somehow related to normal behavior. If we accept the notion that "fantasy is driven by need" (39, p. 118), it can be predicted from any of these points of view that there would be no difference in the content of fantasy productions of individuals of different diagnostic categories since all individuals have the same needs, whether they are psychotic, neurotic or normal. And these are indeed our findings.

If any one thing differentiates our subjects it seems to be the factor of hospitalization, which has been discussed above in the consideration of emotional tone and outcome. The effect of the external environment on the content of fantasy productions has been pointed out by some investigators.

Bumke and Kant (12), for example, report that patients in continuous hydrotherapy hallucinate lobsters and fish. Potter (43) found environmetal factors, in this case a hospital stress situation, affected the fantasy of women awaiting gynecological surgery and obstetrics so that their Rorschach protocols were more similar to each other than to a matched group of clerical workers not in a hospital situation. This "levelling" effect, which also makes our hospitalized patients, regardless of clinical classification, more similar to each other than non-hospitalized subjects, is further evidence of the influence of situational factors, e.g., limitation of environment, on fantasy productions. Bondy (10) in a discussion of the effect of internment camps on personality organization, attributes to the social isolation the eventual similarities in personality which emerge. "In consequence of this isolation in the internment camps, all social veneer and superficial personality traits soon disappear. With startling rapidity the internees reach a common denominator and lose their individuality" (10, p. 462). Bettelheim (9) found significant differences when comparing old and new prisoners and the differences "seemed to originate in personality changes brought about by the impact of camp experiences on the prisoners" (9, p. 437). The unnatural, restricted, and routinized environment to which our hospitalized patients are exposed may easily have a similar kind of effect on their fantasy productions.

F. LEVEL OF INTERPRETATION

Since the analysis of the TAT by content has failed to discriminate between diagnostic classifications, and since it has been suggested by some (32, 54) that analysis of the formal aspects of the productions would be more fruitful, the writer, in the course of analyzing the stories for themes, noted also when there were deviations from the instructions to make up a narrative. Of 16 different kinds of deviation from the instructions,[6] these groups do not make up a homogeneous population in five (Table 13). The specific inter-group differences significant at or beyond the .05 level of confidence are noted in the latter part of Table 10; and there are 29 such differences, which is more than could be expected by chance. The hospitalized psychoneurotics reject more cards than do all other subjects, although each of the hospitalized groups rejects significantly more than each of the non-hospitalized groups. There are only two rejections among the non-hospitalized psychoneurotics, one in one normal group and none in the other. Each of the hospitalized groups also gives significantly more descriptions than each non-hospitalized group. The non-hospitalized subjects seem to be more willing or able to cooperate than the hospitalized ones, since they comply more closely with the directions. All of the clinical groups give significantly more alternate themes than one normal group. (The other normal group also has less of these than any clini-

TABLE 13

LEVELS OF INTERPRETATION FOR WHICH ALL GROUPS OF SUBJECTS DO NOT MAKE UP A HOMOGENEOUS POPULATION

Level	Chi-Square of Homogeneity	p*
III A Symbolic	20.1539	.01
III C Description	27.2000	.001
III I Alternate Themes	22.6152	.001
III J Comments	19.3000	.01
III L Rejection	21.3750	.001

* 5 degrees of freedom.

---

[6] See Appendix B for a description of these deviations.

cal group, but the difference is significant only when comparison is made with the hospitalized psychoneurotic group.) The excess of alternate themes in these groups is no doubt a function of indecision and uncertainty of the subjects in them just as the greater number of questionable outcomes is. Analysis by individuals does not change the picture very much, except that now fewer significant differences appear, but all are in the same direction.

In light of what has been stated in the TAT literature about the formal characteristics of stories of various clinical groups, it is appropriate to see how our data corroborate or are at variance with what has been reported. The fact that both groups of normal subjects give more symbolic stories[7] than each of the clinical groups, significantly more than the psychoneurotics, contradicts reports that this is a characteristic of the stories of schizophrenics (44) and psychotic children (32). The normals also give more abstract stories,[7] although this difference is not significant at the .05 level. It has been said that ambitendency in story trend is also a schizophrenic characteristic (46), yet as noted above this seems to be more characteristic of psychoneurotics. As compared with normals, the schizophrenics do give more such themes, but they seem to be representative of all clinical groups and not distinctive of schizophrenics. Story continuations from one picture to the next has been listed as diagnostic of schizophrenics (44), but again our data show this to be more characteristic of normal stories (Table 11). Statements that the depicted situations are unreal are also more frequent among normals; and the difference is significant between one such group and the hospitalized general neuropsychiatric group. The schizophrenics and the general neuropsychiatric patients make significantly more comments about the picture than do both of the psychoneurotic groups and one of the normal groups. This too has been called a psychoneurotic characteristic (46).

Other unusual kinds of interpretation are also more common among normals and often do not even appear among some of the clinical groups, as shown in the latter part of Table 11. For example, having the central character outside the picture is an interpretation given by six normal subjects and two schizophrenics; and three normals included the examiner in the story,[8] whereas only one each among the hospitalized psychoneurotics and the general hospital group did likewise. This latter has been called an exclusively psychotic characteristic (32). As would be expected, peculiar verbalizations, confused stories, and stories with no conceivable connection to the picture are significantly more characteristic of schizophrenics than other subjects, although the actual number of schizophrenics who produce stories with such characteristics are very few.

The finding that there is more differentiation among groups on the basis of the formal characteristics of the stories than on the basis of content is comparable to the results of Anastasi and Foley (3). However, just as they report, none of the items occurs with a sufficiently high frequency in any of the groups to be regarded as a regular feature of its stories. Each of the deviations in the check list was observed in only a small minority of each population and very few of the items were absent from any group.

---

[7] An example of a symbolic story contributed by a normal subject is as follows: It looks like Dante's inferno. Some sort of monstrosity—a primeval picture—symbolic of evil. This creature here is exploring the unknown—symbolic of man. Might even say it's all a picture of life—dark, cloudy, devastation. The way it's going to turn out is that this small creature is going to defeat this evil up here. More or less a symbolic approach. (Card 11)

An example of an abstract story contributed by a normal subject is as follows: It looks like Salvador Dali. Well, this is an emotional picture. It shows happiness and contentment of this couple despite storms and adversities. They are secure in each other's love. Happiness is expressed here—the ultimate in marital bliss—peace and contentment. (Card 19)

[8] One of the characters in each of these stories was a "Dr. Eron."

IV. NORMATIVE DATA FOR THE TAT

It is obvious that the TAT pictures themselves call forth certain kinds of stories with characteristic emotional tone and that they are as important as the clinical classification of the particular subject in determining what kind of a response he will make. The identification of the most popular responses to each card has been designated a study of "picture pull." It has been felt that each individual picture has its own stimulus properties which evoke themes, identifications, feelings, etc. which are peculiar to it and which differ from those elicited by other pictures. Appendix C lists for each picture the most frequent themes contributed by all subjects, classified by group. Also appearing in Appendix C are the frequency ratings of emotional tone and outcome, frequency of shift, type and frequency of all perceptual distortions, unusual attention to details, and unusual interpretations. Information of this type is indispensable for anyone who wishes to make a valid interpretation of a TAT protocol. It also gives a good indication of what cards should be used when there is insufficient time to administer the whole set of cards and the definite area of adjustment is known about which information is desired.

Certain pictures have been said to be more stimulating than others (31, 46). An analysis of the number of themes elicited by each card corroborates this claim. Table 14 lists the pictures in the order of their stimulatory value for these 150 subjects. There was only one significant intergroup difference in the number of themes aroused by each picture. This was for Card 16, in which Group (C2) has significantly more themes than Group (GNPH) (chi square 6.1538, significant at .05 level of confidence).

Since the test is usually administered in two parts with one half of the cards presented in each session, it is desirable to know if there is any difference in the kind of story told for either half. All significant differences in themes and levels of interpretation for the first and second 10 pictures appear in Table 15, and there are significant differences for 48 themes and eight levels, both many more than could be expected by chance. Themes and levels for which there is no occurrence in one or the other parts of the test are listed in Table 16. It is obvious from these two tables that each half of the test has a differential effect on the kind of stories it elicits. A study by Kannenberg (29) indicates that this is not a function of order of presentation or special instructions to make the stories as imaginative as possible, as had previously been assumed (16).

TABLE 14

Rank Order of Pictures on Basis of Number of Themes Which Each Elicits

Stimulatory Value	Picture	No. of Themes
1	13MF	456
2	20	421
3	18BM	413
4	6BM	395
5	3BM	373
6	4	361
7	12M	352
8	15	340
9	7BM	316
10.5	10	301
10.5	17BM	301
12	8BM	287
13	9BM	273
14	14	264
15	5	255
16	2	239
17	1	237
18	19	225
19	11	202
20	16	199

TABLE 15

DIFFERENCES IN NUMBER OF THEMES IN PICTURES 1–10 AND 11–20 SIGNIFICANT AT OR BEYOND .05 CONFIDENCE LEVEL

Theme		1–10	11–20	Chi-square	P
No.	Name				
IA1a	Pressure from Parents	260	29	183.0450	.001
IA1b	Succorance from Parents	172	14	132.5215	.001
IA1c	Nurturance to Parents	20	5	7.8400	.005
IA1k	Disappointments to Parents	25	5	12.0333	.001
IA1f	Departure from Parents	73	11	44.2976	.001
IA1g	Concern of Parents	30	3	20.4848	.001
IA1i	Death or Illness of Parents	79	16	41.7789	.001
IA1m	Filial Obligation	38	3	28.1951	.001
IA1p	Marriage of Child	34	1	29.2571	.001
	Total Parent Disequilibrium	809	135	479.7976	.001
IA2a	Pressure from Partner	102	24	47.0555	.001
IA2b	Succorance from Partner	38	7	20.5000	.001
IA2c	Nurturance to Partner	25	8	7.7575	.01
IA2e	Aggression towards Partner	8	35	15.7209	.001
IA2f	Departure from Partner	48	7	29.0909	.001
IA2i	Illicit Sex with Violence	1	16	11.5294	.001
IA2j	Death or Illness of Partner	18	44	10.0806	.005
IA2n	Competition by Partners for Hero	34	4	22.1316	.001
IA2s	Unrequited Love	5	19	7.0417	.01
IA2r	Seduction by Partner	17	3	8.4500	.005
	Total Sibling Disequilibrium	29	12	6.2439	.015
IA3a	Pressure from Peer	8	44	23.5577	.001
IA3b	Succorance from Peer	14	120	82.2761	.001
IA3e	Aggression towards Peer	23	60	15.6144	.001
IA3d	Aggression from Peer	10	49	24.4746	.001
IA3i	Death or Illness of Peer	16	41	10.1052	.005
IA3j	Belongingness	59	29	9.5568	.005
IA3k	Unappreciated by Peers	1	8	4.0000	.045
IA3l	Competition from Peers	3	30	20.4848	.001
	Total Peer Disequilibrium	154	482	168.1273	.001
	Total Interpersonal Disequilibrium	1396	1021	57.8718	.001
IB1	Aspiration	196	48	83.8073	.001
IB4	Behavior Disorder	56	112	18.0059	.001
IB6	Moral Struggle	25	5	12.0333	.001
IB7	Guilt—Remorse	44	99	20.2500	.001
IB8	Drunkenness	26	106	47.2803	.001
IB11	Occupational Dissatisfaction	129	28	63.6943	.001
IB13	Retribution	23	48	8.1127	.005
IB16	Religion	14	112	74.6746	.001
IB17	Loneliness	16	51	17.2537	.001
IB20	Vacillation	34	132	56.6807	.001
	Total Intrapersonal Disequilibrium	972	1186	21.0236	.001
IC2	Legal Restriction	27	61	12.4842	.001
IC5	Aggression from Environment	49	218	107.2283	.001
IC6	Escape from Perilous Environment	6	69	52.0324	.001
	Total Impersonal Disequilibrium	335	562	132.8426	.001
IIA1f	Contentment with Parents	10	43	19.3207	.001
IIA2c	Contentment with Partner	75	17	35.3152	.001
IIA2d	Reunion with Partner	22	2	15.0417	.001
	Total Partner Equilibrium	111	29	46.8642	.001
IIA3d	Peer Approbation	1	17	14.2223	.001
	Total Peer Equilibrium	13	80	46.8387	.001
IIB1	Self Esteem	2	41	33.5814	.001
IIB3	Reminiscence, happy	14	37	9.4902	.005
IIB4	Retirement	110	36	23.9427	.001
IIC1	Favorable Environment	4	34	22.1316	.001
	Total Impersonal Equilibrium	4	36	24.0250	.001
	Total Equilibrium	334	404	6.4512	.015

(*Continued on next page*)

Theme No.	Theme Name	1–10	11–20	Chi-square	*P*
	Level				
III A	Symbolic	16	72	39.5852	.001
III B	Abstract	2	19	14.0870	.001
III D	Unreal	8	68	53.4405	.001
III G	Autobiographical	31	78	42.3500	.001
III H	Continuations	3	14	5.8823	.02
III I	Alternate Themes	52	123	66.6475	.001
III J	Comments	22	42	5.6406	.02
III L	Rejection	13	33	7.8478	.005

TABLE 16

Themes and Interpretation Levels for Which There Is No Frequency in One Half of the Test

Theme No.	Theme Name	1–10	11–20
IA1h	Incest		2
IA1n	Confession to Parent	18	
IA1o	Bad News to Parent	18	
IA1q	Collusion with Parent	5	
IA4b	Succorance from Sibling	1	
IA4e	Aggression to Sibling	4	
IA3n	Envy of Peer		4
IA3m	Hypnotism		61
IB18	Compensation		14
IB23	Inconsequential Sadness	3	
IB24	Hurt Feelings	18	
IB25	Jealousy (Unspecified)	3	
IB28	Homesickness		4
IIA1b	Resignation of Parent	1	
IIA1e	Fulfillment to Parent	4	
IIA2a	Admiration of Partner	3	
IIA3c	Reunion with Friends		2
IIA3e	Exhibition		36
IIB6	Resignation to Lot	3	
IIB7	Ordinary Activity		42
IIC2	Rescue		2
	Level of Interpretation		
III E	Fairy Tale		10
III F	Central Character Out		8

V. SUMMARY AND CONCLUSIONS

A. Approximately 3,000 Thematic Apperception Test (TAT) stories contributed by 150 male veterans including 50 college students, 25 non-hospitalized psychoneurotics, 25 hospitalized psychoneurotics, 25 hospitalized schizophrenics and 25 miscellaneous neuropsychiatric patients, were analyzed for themes, level of interpretation, emotional tone, outcome, identification of characters, perceptual distortions and attention to details. The data were arranged (*a*) according to total number of responses for all 20 cards presented, and also (*b*) according to number of individual subjects contributing a certain number or more of a given response. Significant differences between groups were calculated on both of these bases by the chi square method, and all differences significant at or beyond the .05 level of confidence were noted. In all, 400 significant inter-group differences were noted, as shown in Table 17. As seen in this table, the two most similar groups are (C1) and (C2), the two normal ones, since there were only 6 significant differences found between them, which is much less than would normally be expected by chance. Next in order of similarity are the normal groups (C1) and the non-hospitalized psychoneurotics (NHPN), and groups (C2) and (NHPN), with 10 and 11 significant differences respectively. Thus the three non-hospitalized groups are the most similar. Next in order of similarity are (HPN) and (HS), the hospitalized psychoneurotics and schizophrenics; (HS) and (GNPH), the hospitalized schizophrenics and the general neuropsychiatric hospitalized group; and (HPN) and (GNPH) each with 12, 16, and 17 significant differences respectively. The groups with the number of significant differences between them closest to this are (NHPN) and (HS), with 25 such differences; all other "cross-comparisons" (i.e., comparisons between hospitalized and non-hospitalized groups) yield a *larger* number of significant differences—up to 46 for group (C1) vs. (GNPH). It would seem, then, that *our six groups of subjects arrange themselves in two clusters according to whether or not they are hospitalized,* since the three non-hospitalized groups are most similar, followed by the three hospital groups, and finally by comparisons between groups of the two different classes.

B. When each of the six groups was compared with every other group on each of the six variables specified below, not many more significant differences emerged than would be expected by chance. The findings can be summarized briefly as follows:

1. *Emotional tone.* The most frequent stories of our entire group are moderately sad regardless of the clinical classification of the subjects contributing them. The hospitalized subjects more frequently tell stories which are neutral in emotional tone than do the non-hospitalized subjects. This appears to be more a function of hospitalization than of seriousness of illness, because there is no steady progression in the number of such stories from normals through psychoneurotics to psychotics.

2. *Outcome.* The conclusions offered by the subjects seem to be more a projection of their own creativity than is the emotional tone, since there is considerably more inter-individual and inter-group variability for this dimension. However, the most frequent outcome is *happy* for all except the schizophrenic group, which contributes more neutral outcomes. As in emotional tone the hospitalized subjects give less extreme outcomes, either very happy or very sad, than do the non-hospitalized subjects; and again the hospitalization is considered the chief factor in accounting for this greater blandness.

3. *Shift.* Slightly less than half of the sub-

TABLE 17

NUMBER OF SIGNIFICANT DIFFERENCES BETWEEN GROUPS

Groups	Themes	Levels	Tone	Outcome	Shift	Distortions	Total	Grand Total	Rank Similarity
A. Total responses									
C1-C2	2	1		1			4		
C1-NHPN	1	2		2	1		6		
C1-HPN	16	5	3	5			29		
C1-HS	12	4	3	6	5		30		
C1-GNPH	14	5	5	6	3	1	34		
C2-NHPN	4	2		3	1		10		
C2-HPN	10	7	2	3			22		
C2-HS	9	4	3	3	2		21		
C2-GNPH	21	7	3	4	1	1	37		
NHPN-HPN	8	3	3	5	1		20		
NHPN-HS	4	6	2	2	4		18		
NHPN-GNPH	10	4	3	5	3	1	26		
HPN-HS	5	2		2	1		10		
HPN-GNPH	14	1				1	16		
HS-GNPH	11	1			1	1	14		
Total							297		
B. Individuals									
C1-C2									
C1-NHPN		1					1		
C1-HPN	3	2	1	2			8		
C1-HS	3	1		1			5		
C1-GNPH	2		2	4		1	9		
C2-NHPN	1						1		
C2-HPN	1	2		2			5		
C2-HS	1	1	1				3		
C2-GNPH	2					1	3		
NHPN-HPN			1	1			2		
NHPN-HS			1				1		
NHPN-GNPH		1	3	1		1	6		
HPN-HS									
HPN-GNPH						1	1		
HS-GNPH									
Total							45		
C. Card by card						Details			
C1-C2				1		1	2	6	1
C1-NHPN	1			1		1	3	10	2
C1-HPN	4			3			7	44	13.5
C1-HS	4			1			5	40	11.5
C1-GNPH	1			1	1		3	46	15
C2-NHPN								11	3
C2-HPN	3	1		2		1	7	34	10
C2-HS	2			2		1	5	29	9
C2-GNPH	3			1			4	44	13.5
NHPN-HPN	1		2	2			5	27	8
NHPN-HS	1		1	2	1		5	24	7
NHPN-GNPH	2		1	3	1	1	8	40	11.5
HPN-HS	2						2	12	4
HPN-GNPH								17	6
HS-GNPH	1					1	2	16	5
Total							58		
Grand total								400	

jects terminate their stories with conclusions which are happier than the general tone of the body of the story. A shift to happier conclusions is less characteristic of our schizophrenic subjects than it is of the normal ones.

4. *Perceptual Distortions.* Misidentification of the sex of characters in the pictures, and other perceptual distortions, are frequent, and no more characteristic of any of our specific nosological groups than of our normal subjects. It was demonstrated that to use misidentification of the sex of the figure in Card 3 BM as an index of a high feminine component in the male personality is fallacious.

5. *Themes.* When the individual groups are compared on the basis of our 125 theme check list only a few more differences appear than could be expected by chance. Themes of aggression, hostility, frustration, restriction, death, violence, and guilt are common to all groups. For no given theme does any *one* group of subjects differ from *all* others, and *no* group differs from *both* normal groups in more than five themes. The subjects were rearranged into larger groups, and comparisons were made between all normal and all psychoneurotic subjects, and between all hospitalized and all non-hospitalized subjects. On this basis more significant differences emerged than could be explained by chance. The nature of these differences, as well as of those among the individual groups, was discussed both in the light of the present literature on the TAT and possible personality dynamics.

6. *Level of Interpretation.* In an attempt to see if more inter-group differentiations could be made on the basis of a formal, rather than a content, analysis, the stories were classified according to deviations from the instructions to make up narratives; and many more significant differences were obtained than could be expected by chance. Again our results were compared with those reported in the literature; and many of the differences which have been reported, when put to the statistical test, were found not only to be non-significant, but often reversed.

C. We interpret these findings to mean the following:

(1) Fantasy productions of various clinical groups do not differ significantly from each other or from the productions of a normal population.

(2) Fantasy productions are affected by the external "objective" environment of the individual, in this case a limiting, routinized, hospital environment.

(3) The TAT, as a fantasy technique, should not be used as a diagnostic instrument in the sense of yielding "signs" or "patterns" characteristic of specific nosological entities.

D. In making a valid interpretation of a given protocol, due consideration must be given to the stimulus properties of the cards themselves, which appear to be as important a factor in determining an individual's response as the actual clinical group in which he may have been classified. As an aid to the proper utilization of this technique, a table has been prepared and is appended (Appendix C) which gives for each of the 20 cards in the adult, male series of the Third Edition of the TAT, the most popular responses contributed by 150 subjects in terms of emotional tone, outcome, perceptual distortions, identification, attention to details, and thematic content.

E. Although the findings negate much of what has been reported in the TAT literature, they are not to be construed as an argument against the validity of the instrument. The TAT was introduced to investigate the fantasy of normal individuals. It was assumed that from this fantasy there could be inferred important strivings and needs in the narrator. Nothing here reported contradicts this assumption. What is demonstrated is the lack of justification for using the method as a diagnostic device for the differentiation of nosological groups. Schizophrenics, psychoneurotics, and normal persons all have the same needs and strivings, preoccupations and attitudes, delusions and misconceptions. There may be a difference in emphasis and degree from individual to individual, but not from group to group. Normal persons check up on their ideas, and correct their delusional thinking through techniques of social participation, role taking, and the sharing of perspective, more readily than individuals with behavior disorders, in whom these skills are not as well practiced. But the fact remains, according to the evidence of the present study, that we all fantasy much the same things. Therefore, a technique such as the TAT cannot be used as a diagnostic instrument in the sense of pigeon-holing people into different Kraepelinian categories. If we mean by diagnosis the determination of the dynamics and content of an individual personality and the understanding of his preoccupations and conceptions, the TAT may well be of great value. It is a diagnostic instrument in the sense that it gives an understanding of the individual in his own life-setting. The results of this study indicate that it does not yield differential patterns for classification into Kraepelinian nosological entities.

APPENDIX A

RATING SCALES FOR EMOTIONAL TONE OF *TAT* STORIES

Key for all scales
—2 Very sad
—1 Sad
0 Neutral
+1 Happy
+2 Very happy

General Rating Scale for Emotional Tone of Stories:

—2 Complete failure, submission to fate, death, murder, suicide, illicit sex with violence, revenge, aggressive hostility, severe guilt, complete hopelessness.

—1 Conflict with attempt at adjustment, rebellion, fear, worry, departure, regret, illness, physical exhaustion, resignation toward death, loneliness.

0 Description, lack of affect, balance of positive and negative feelings, routine activities, impersonal reflection.

+1 Aspiration; desire for success and doubt about outcome, compensation for limited endowment. Description with cheerful feeling, reunion with friends, contentment with world, feeling of security.

+2 Justifiably high aspiration. Complete satisfaction and happiness. Reunion with loved ones.

? Can't make up a story.

General Rating Scale for Outcomes:

—2 Complete failure, submission to fate, death, murder, suicide, extreme punishment, extreme remorse.

—1 Some frustration: incomplete success in attaining goal, goal attained at expense of happiness, disappointment to friends and family, acceptance of unsatisfactory situation or submission to authority.

0 Continuation of ordinary situation, balance of happy and unhappy situations.

+1 Moderate success, reunion with friends, recovery from temporary disability or depression, happiness in success of others.

+2 Great success, discovery, and/or happiness. Extreme contentment, marital bliss, unusual good fortune, reunion with loved ones.

? Cannot give an outcome, even when explicitly asked for, conditional (if) outcomes, alternative outcomes of different emotional value.

Rating Scales for Emotional Tone of Individual Cards:

(1)

—2 Complete frustration and hopelessness with no resistance.

—1 Dejected, inadequacy with attempt to adjust.

0 Frustration with no depression. Aspirations balanced by conflict. Lack of feeling tone.

+1 High aspirations with cooperation but some hindrance.

+2 High aspiration with approbation and no conflict.

(2)

—1 Conflict between ambition and duty to family with some feelings of guilt or apprehension, disappointment with lot, jealousy.

0 Description of picture, planning for future with no apparent affect.

+1 Future planning, description with cheerful feeling.

+2 Complete satisfaction with present status and life's accomplishments.

(3BM)

—2 Uncontrolled emotionality, murder, mentally ill, complete frustration, death of loved one, suicide.

—1 Self-pity, aggressive parental pressure, transitory depression, adolescent confusion, physical incapacity.

(4)

—2 Desire for revenge, murder, aggressive hostility.

—1 Disillusionment, occupational failure, conflict over extra-marital relations, jealousy, pressure from mate.

(5)

—2 Overdominant parent or mate, extreme parental or partner pressure, aggressive hostility, murder.

—1 Parental or partner pressure without aggression, loneliness, slight frustration.

0 Checking on room or occupants, doing household duties.

+1 Unexpected gift, welcome guest, good news.

(6BM)

—2 Death, bad news, severe guilt, conflict over social acceptance of sexual role.

—1 Parental pressure, filial obligation, conflict over desires and duties, departure from parental home.

(7BM)

—2 Disappointment of parents in child, guilt, repeated failure.

—1 Disagreement, recalcitrance, rebellion against parental authority, feelings of inadequacy.

0 Parental advice, impersonal discussion, counselling.

+1 Aspiration with encouragement and/or advice.

(8BM)

−2 Murder, death, extreme guilt.
−1 Worry and concern about accident and operation, frustration of ambitions.
0 Descriptive, perfunctory, lack of emotional involvement, impersonal reflection.
+1 Aspiration, hope and planning for future. Adventurous daydreaming.

(9BM)

−1 Economic misfortune, physical exhaustion, danger of combat, social disapproval.
0 Men at rest, pure description, no emotional involvement.
+1 Comradely feeling, contentment, carefree, happy-go-lucky, lack of concern for convention.

(10)

−2 Death, extreme sorrow, tragedy.
−1 Departure, leaving loved ones, personal failure, being comforted for minor misfortune.
0 Lack of affect, balance of conflict.
+1 Reunion, happiness, acceptance, feelings of pleasure.
+2 Marital bliss, extreme contentment, satisfaction and good adjustment.

(11)

−2 Life is futile, horrible, complete absence of hope, no avoidance of fate. Death, destruction, war.
−1 Struggle against aggressive forces, animals fighting, story detached from reality.
0 No emotional involvement, little interpersonal action, description.
+1 Vacation, pleasure trip, happy people.

(12M)

−2 Death, suicide, malpractice (hypnotism) with aggression, rape, curse.
−1 Reconciliation to death; illness, parental pressure.
0 Hypnosis with no harm involved (experimentation with or demonstration of hypnosis). Being awakened from sleep.
+1 Reunion.

(13MF)

−2 Illicit sex with violence, rape, death, murder for infidelity.
−1 Disillusionment with sexual experience, regret for illicit sex. Illness of wife.

(14)

−1 Resignation to death of relative. Reflection on worldly conflicts, with or without appeal to religion, loneliness.
0 Daydreaming without emotional involvement, any other theme with no emotional involvement, adolescent reverie.
+1 Contentment with environment, appreciation of world around.
+2 Happy, well-adjusted hero, vacation, planning for happiness.

(15)

−2 Death of close relative, loneliness for deceased, mourning, hopelessness, hero rejected by society, suicide.
−1 Impersonal speculation on death, return of dead to cemetery, visiting grave of friend.
0 Description of painting or picture, no affect.

(17BM)

−1 Vindictiveness, revenge, trying to escape from unfavorable environment, fear, inadequacy.
0 Vacillation in plot (balance of happy and sad themes), doing routine job of acrobatics.
+1 Compensation for limited endowment, desire for success with uncertainty about outcome.
+2 Hero happy and successful. Display of physical prowess, adulation of crowd, winning of contest.

(18BM)

−2 Suicide, manslaughter, thwarted escape, hallucinations, delusions.
−1 Environmental frustration, accident, ordinary drunkenness, personal sorrow not of serious proportion.
0 Description of poster or painting. No emotional involvement.

(19)

−2 Death due to forces of nature or war.
−1 Fear (child's fear of supernatural), bad storm with little or no emphasis on comfort of home.
0 Description of picture.
+1 Comfort of home during storm, feeling of security.

(20)

−2 Death of loved one, suicide.
−1 Disappointment in love, worry, feeling of rejection, economic pressure, disillusionment, loneliness.
0 Out for a walk, description, no feeling tone.

APPENDIX B

CHECK LIST OF THEMES AND INTERPRETATION LEVELS

- I. Disequilibrium (tension)
  - A. Interpersonal
    - 1. Parent (includes parent substitutes such as older authority figures, etc.)
      - a. *Pressure*—parent or parent figures are prohibitive, compelling, censuring, punishing, disapproving, interfering, checking up, disagreeing with, quarreling with, restraining or unduly influencing child.
      - b. *Succorance*—child seeks or receives aid, help, advice, consolation, from parent.
      - c. *Nurturance*—child bestows or offers aid, advice, consolation to the parent.
      - d. *Aggression from*—physical harm inflicted or intended upon child by parent.
      - e. *Aggression to*—physical harm inflicted or intended upon parent by child.
      - f. *Departure*—child is taking leave of parental home, is separated from parents.
      - g. *Concern*—parent is worried over physical or mental well-being of child.
      - h. *Incest*—actual or contemplated.
      - i. *Death or illness* of parent.
      - j. *Death or illness* of child.
      - k. *Disappointment to*—parent is disappointed in child's behavior or accomplishments.
      - l. *Disappointment in*—child is disappointed by parent.
      - m. *Filial obligation*—child feels it is his duty to remain with, comply with, or support parents.
      - n. *Confession*—child tells parent of some misdemeanor or crime he has committed.
      - o. *Bad news*—child brings bad tidings to parent (e.g., death of other parent, brother, etc.) or friend brings tidings of death of son.
      - p. *Marriage*—child tells parent of past or impending marriage, parent objects to marriage.
      - q. *Collusion*—parent and child are planning or executing together some anti-social act.
      - r. *Parental conflict*—child is concerned over marital problems of parents.
      - s. *Lost*—child has been abandoned by parents or is lost.
    - 2. Partner (wife, sweetheart, opposite-sexed peer).
      - a. *Pressure*—partner is prohibitive, compelling, censuring, punishing, quarreling with, etc.
      - b. *Succorance*—hero receives or seeks aid, comfort, consolation, assistance from partner.
      - c. *Nurturance*—hero bestows or offers to partner, aid, consolation, advice, etc.
      - d. (*Aggression from*—physical harm inflicted upon or intended for hero by partner.)*
      - e. *Aggression to*—physical harm inflicted upon or intended for partner.
      - f. *Departure from*—hero is leaving partner either temporarily or permanently or is separated from partner.
      - g. *Concern*—hero is worried over physical or mental well-being of partner.
      - h. *Illicit sex*—extra- or pre-marital heterosexual relationship, non-incestuous, includes "petting."
      - i. *Illicit sex with violence*—rape.
      - j. *Death or illness* of partner.
      - k. *Disappointment to*—partner is disappointed in hero's behavior or accomplishments.
      - l. (*Disappointment by*—hero is disappointed by partner).
      - m. *Jealousy*—hero is jealous of partner's attention to others.
      - n. *Competition*—hero is object of competition between two or more admirers.
      - o. *Cuckold*—hero discovers partner has been having extra-marital relations or finds her raped.
      - p. *Decision*—hero must choose between marriage and not or between two partners.
      - q. *Pursuit*—hero is wooing or trying to get partner to submit.
      - r. *Seduction*—hero is being talked into relationship or is being actively pursued.
      - s. *Unrequited*—hero's love is unreturned, impotence.
      - t. *Restraint*—hero and partner can't be married or have intercourse (because of economics, deformities, menstrual periods, etc.)
      - u. *Childbirth*
    - 3. Peer (like-sexed, approximately same age).
      - a. *Pressure*—like-sexed peers are prohibitive, compelling, censuring, punishing, disapproving, interfering, restraining, etc.
      - b. *Succorance*—hero seeks or receives aid,

---

* Themes in parentheses have not been found in original sample of 150 protocols but are included for completeness.

advice, consolation, from peer.
- c. (*Nurturance*—hero bestows or offers aid, advice, consolation upon peer).
- d. *Aggression from*—physical harm inflicted upon or intended for hero by peer.
- e. *Aggression to*—physical harm inflicted upon or intended for peer by hero.
- f. *Desertion*—hero has been deserted by his companions or is lost from them, or is departing from them.
- g. *Concern*—hero is worried about physical or mental well being of his friends.
- h. *Homosexuality*
- i. *Death or illness* of friend.
- j. *Belongingness*—desire expressed to be with or accepted by peers.
- k. *Unappreciated* by peers.
- l. *Competition*—hero is competing with peer in game or contest.
- m. *Hypnotism*
- n. *Envy*—hero is envious of peer's accomplishments, capacities, etc.
- o. *Revenge*—hero is anxious to exact revenge from a peer for some past deed or is taking the revenge.
- p. *Disregard for*—hero is unconcerned over welfare of his peers.
- q. *Bad influence*—hero's friends are considered unfavorable associates by him, his family, or narrator.

4. Sibling
    - a. *Pressure*—sibling is prohibitive, compelling, censuring, punishing, disapproving, interfering, restraining, etc.
    - b. *Succorance*—hero seeks or receives aid, advice, consolation from sibling.
    - c. (*Nurturance*—hero bestows or offers aid, advice, consolation upon sibling).
    - d. (*Aggression from*—physical harm inflicted upon or intended for hero by sibling).
    - e. *Aggression to*—physical harm inflicted upon or intended for sibling by hero.
    - f. *Rivalry*—hero and sibling are competitors for same objective.
    - g. (*Concern*—hero is worried about physical or mental well-being of his sibling).
    - h. *Incest*—with sibling.
    - i. *Death or illness* of sibling.

B. Intrapersonal
1. *Aspiration*—dreaming of future, hoping for future, determination.
2. *Inadequacy*—realization, whether justified or not, of lack of success; individual is at a loss to cope with situation.
3. *Curiosity*—wondering about construction of object, contents of room, etc.; desire to observe, inquire, explore, investigate; to acquire facts.
4. *Behavior disorder*—personal maladjustment of all sorts, neurosis, psychosis, sleeplessness, hallucination, withdrawal, eccentricity, morbid preoccupation.
5. *Suicide*—attempted or completed, preoccupation with.
6. *Moral struggle*—concern over what is right and wrong, hesitancy in indulging in some act because of ethical prescriptions.
7. *Guilt*—remorse.
8. *Drunkenness*
9. *Fear*—hero is apprehensive, alarmed, terrified, of some person, thing, force.
10. *Rumination*—inability to understand world and its problems, adolescent reverie, man's insignificance, "what am I?" "what's it all about?"
11. *Occupational concern*—deciding between jobs, considering vocations, dissatisfied with present employment, fails at present employment.
12. *Physical illness or death of central character*—(other than when in a peer, partner, sibling, parent relationship, which are included in those specific categories) other than suicide.
13. *Retribution*—forced to atone or be punished for some antisocial act.
14. *Reminiscence, sad*—individual is unhappy in his memories of the past or contemplation of future.
15. *Intra-aggression*—does physical harm to self, short of suicide (which is carried under that heading), either thru accident or deliberate.
16. *Religion*—prayer, seeking consolation from God, religious conflict, religious awakening.
17. *Loneliness*—central character misses someone, is an outcast, friendless, homeless.
18. *Compensation*—when individual has one characteristic or stroke of fortune to make up for another bad characteristic or misfortune.
19. *Vacillation*—wasting time, putting off a distasteful task, procrastination, loitering.
20. *Acquisition*—desire expressed by central character to acquire material things, or is working greedily for possession of goods; miserliness.
21. *Exhaustion* — characters completely "pooped" from overexertion.
22. *Revenge*—hero is preoccupied with wish to retaliate for some past wrong from unspecified individual (when peer, carried under that heading).
23. *Sad* over death of dog, broken toy, loss of object, etc.

24. *"Hurt feelings"*–central character is very sensitive over some happening, slight, insult, etc.
25. *Jealousy*–of unspecified individual (if specific person named, carried under appropriate heading).
26. *Self-pity*–hero feels sorry for self.
27. *Resurrection*–return of the dead, supernatural.
28. *Homesickness*–expressed by central character.
29. *Grief*–over loss of some unspecified individual.

C. Impersonal
  1. *Economic pressure*–individual is compelled to, or prohibited from, or limited in, doing something because of the lack of money.
  2. *Legal restriction*–individual is incarcerated, arrested or detained against his will.
  3. *Generalized restriction*–environment is generally frustrating because of backwardness, danger, lack of opportunity, etc.; hopelessness of life.
  4. *Aggression towards environment*–robbery, accident, murder of unspecified individual (if individual named, carried under the proper heading).
  5. *Aggression from*–impersonal source, accident, animal, nature, disease, etc.
  6. *War*
  7. *Escape from perilous environment*–individual is in the act of getting out (doesn't include rescue, which comes under II).

II. Equilibrium (adjustment)
  A. Interpersonal
    1. Parent
      a. *Cooperation*–parent is working with child towards his own goal.
      b. *Resignation*–parent is resigned to child's activity.
      c. *Idealization*–child idealizes parent.
      d. *Reunion*–of child with parents.
      e. *Fulfillment*–child lives up to expectations of parents.
      f. *Contentment*–blissful home situation.
      g. *Ordinary familial activity*–just talking, spending an evening, etc.
    2. Partner
      a. *Admiration*
      b. *Cooperation*–between mates, so that both are working for same end and are happy.
      c. *Contentment*–mutual love and affection.
      d. *Reunion*
      e. *Ordinary activity*
    3. Peer
      a. *Cooperation*–working together for same goal.
      b. *Congeniality*–sociability, friendliness, satisfactory affiliation.
      c. *Reunion*–with friends.
      d. *Approbation*–hero is lauded, appreciated, by friends and peers.
      e. *Exhibition*–hero is amusing, entertaining, attracting others.
      f. *Ordinary activity*–everyday association with nothing particular happening.
    4. Sibling (no II themes found in original sample of 150 cases).
  B. Intrapersonal
    1. *Self esteem*, confidence, belief in own superiority.
    2. *Tranquility*–peace of mind, content with environment and own accomplishments, enjoyment, aesthetic appreciation.
    3. *Reminiscence, happy*–individual is content with his memories.
    4. *Retirement*–central character is asleep, resting, no physical exhaustion.
    5. *Occupational satisfaction*
    6. *Resignation to lot*–individual has resigned himself to whatever situation.
    7. *Ordinary activity*–individual is going about his everyday activities.
  C. Impersonal
    1. *Favorable environment*–individual is being helped by favorable circumstances, is enjoying his surroundings.
    2. *Rescue*–successful escape from perilous environment or cessation of noxious stimuli (already accomplished).

III. Level of interpretation
  A. *Symbolic*–depiciting an idea or moral.
  B. *Abstract*–depicting a feeling.
  C. *Descriptive*–no action depicited, no story, mere description of picture.
  D. *Unreal*–"This is a picture," "This is someone's dream," etc.–can't accept as a real situation.
  E. *Fairy tale*–legend, impossible happening.
  F. *Central character is not in picture.*
  G. *Autobiographical*–where narrator is character in the story, "like what happened to me," inserts reference to self.
  H. *Continuations*–refers to past stories.
  I. *Alternate themes* given for same picture.
  J. *Comments* about artistic merits of picture, recognition of artist or painting.
  K. *Denial of a theme*–"This is not–" when it is commonly given.
  L. *Rejection*
  M. *Peculiar verbalizations*
  N. *Confused*, no single discernible plot, impossible conclusion, etc.
  O. *Includes examiner* in story.
  P. *Absolutely no connection* of story to picture.
  Q. *Humorous*

APPENDIX C

NORMATIVE DATA FOR THE *TAT*[1]

This appendix includes a card by card tabulation of TAT responses analyzed in terms of the following variables:

A. Emotional tone of stories
B. Outcome of stories
C. Shift in emotional tone from body of story to outcome
D. Most frequent themes (included are all themes contributed by at least 5 subjects)
E. Identifications of characters and objects (for cards on which such identification is not uniform, and five or more subjects produce the divergent identification)
F. Attention to details (for cards on which five or more subjects single out some unusual or minute aspect of the picture for elaboration in their stories).

The population on whom this analysis is based includes 50 normal male college students, 25 non-hospitalized male psychoneurotics, 25 hospitalized male psychoneurotics, 25 hospitalized male schizophrenics and a group of 25 males representing a general neuropsychiatric hospital population. All groups are similar as to age, education, intelligence, and marital status and all subjects are veterans of World War II. For the purposes of presentation in this appendix these 150 subjects have been divided into two groups of 75 each. Group *H* consists of the hospitalized subjects and Group *NH* of the non-hospitalized ones. This division of subjects into two groups is justifiable since the more detailed analysis in the text reveals that hosptalization is the chief significant factor differentiating these subjects. The numbers appearing in the body of the Appendix represent percentages based on a total N of 75 in each group.

Readers desiring these data broken down into the individual groups, with statistically significant differences noted where they occur, may obtain the same from the author at the Department of Psychology, Yale University, 333 Cedar St., New Haven, Conn.

Card 1

A. Emotional Tone Ratings

	−2	−1	0	+1	+2
H	0.0	29.3	53.3	13.3	4.0
NH	1.3	34.7	30.7	24.0	9.0

Most frequent story is neutral to moderately sad.

The hospitalized subjects contribute significantly more neutral stories.

B. Outcome Ratings

	−2	−1	0	+1	+2	?
H	1.3	8.0	37.3	34.7	8.0	8.0
NH	0.0	17.3	30.7	25.3	17.3	8.0

Most frequent outcome is happy.

C. Shift

	Happier	Sadder	No Shift
H	38.7	5.3	56.0
NH	37.3	9.3	53.3

About half of the subjects shift.

Significantly more subjects shift to happier than sad conclusions.

D. Most Frequent Themes[a]

		H	NH
1. IB1[b]	Aspiration	28.0	46.7
2. IA1a	Parental pressure	33.3	41.3
3. IA3j	Belongingness	20.0	25.3
4. IB3	Curiosity	21.3	10.7
5. IB2	Inadequacy	21.3	9.3
6. IB11	Occupational dissatisfaction	6.7	8.0
7. IB19	Vacillation	5.3	5.3
8. IC1	Economic pressure	1.3	5.3
III G	Autobiographical	9.3	4.0

[a] Contributed by five or more subjects and listed in decreasing order of frequency.

[b] Code number to facilitate looking up definition on check list.

Card 2

A. Emotional Tone Ratings

	−2	−1	0	+1	+2
H	0.0	34.7	44.0	21.3	0.0
NH	2.7	53.3	28.0	16.0	0.0

Most frequent story is moderately happy.

Hospitalized subjects contribute more neutral stories (not significant).

B. Outcome Ratings

	−2	−1	0	+1	+2	?
H	0.0	8.0	48.0	46.7	0.0	0.0
NH	2.7	18.7	26.7	45.3	4.0	1.3

Most frequent outcome is happy.

Hospitalized subjects contribute significantly more neutral outcomes.

C. Shift

	Happier	Sadder	No Shift
H	41.3	5.3	53.0
NH	49.3	10.7	40.0

About half of the subjects shift.

Significantly more subjects shift to happier than sad conclusions.

---

[1] Included are all 20 cards (3rd Edition, Harvard University Press) recommended for use with adult males (41).

D. Most Frequent Themes

	H	NH
1. IB11 Occupational Concern	28.0	33.3
2. IB1 Aspiration	26.7	33.3
3. IC1 Economic pressure	17.3	10.7
4. IA1a Parental pressure	8.0	14.7
5. IA1b Succorance from parent	8.0	14.7
6. IIB2 Tranquility	8.0	12.0
7. IA1m Filial obligation	10.7	8.0
8. IA1f Departure from parent	8.0	8.0
9. IA2m Competition for partner	5.3	4.0
10. IA3j Belongingness	2.7	4.0
11. IIC3 Generalized restriction	0.0	6.7
12. IIB5 Occupational satisfaction	4.0	2.7
III C Descriptive	13.3	5.3
III A Symbolic	2.7	10.7
III J Comments about artistry, etc.	5.3	2.7
III D Unreal	4.0	2.7

E. Identifications

	H	NH
1. Father-Mother-Daughter	40.0	29.3
2. Husband-Wife-Unrelated girl	10.7	6.7
*3. Son-Mother-Daughter	4.0	16.0
4. Son-Mother-Unrelated girl	1.3	4.0
5. Brother-Sister-Sister	10.7	5.3
6. Man-two women (all unrelated)	6.7	2.7
7. Two characters only specified	9.3	17.3
8. Four characters specified (includes figure in extreme background	4.0	5.3
9. Only one character specified (girl)	9.3	14.7

F. Attention to Details

	H	NH
1. Pregnancy of women on left	25.3	22.7
2. Books held by young girl	22.7	24.0

* Difference is significant at or beyond .05 level.

Card 3 BM

A. Emotional Tone Ratings

	−2	−1	0	+1	+2
H	52.0	44.0	2.7	1.3	0.0
NH	52.0	46.7	1.3	0.0	0.0

Preponderance of stories are moderately to very sad.

B. Outcome Ratings

	−2	−1	0	+1	+2	?
H	24.0	9.3	18.7	28.0	0.0	20.0
NH	25.3	10.7	21.3	26.7	2.7	13.3

Most frequent outcome is moderately happy.

C. Shift

	Happier	Sadder	No Shift
H	46.7	1.3	52
NH	52.0	2.7	45.3

About one half of all subjects shift to happier conclusions.

Only rare individuals shift to sadder conclusions.

D. Most Frequent Themes

	H	NH
1. IB5 Suicide	26.7	16.0
2. IA1a Pressure from parents	20.0	21.3
3. IB4 Behavior disorder	13.3	20.0
*4. IB7 Guilt–remorse	22.7	8.0
*5. IC4 Aggression to environment	18.7	5.3
6. IA1i Death of parent	9.3	13.3
7. IB12 Physical illness or death of central character	13.3	9.3
8. IC2 Legal restriction	16.0	6.7
9. IB21 Exhaustion	13.3	6.7
10. IB13 Retribution	9.3	9.3
11. IB2 Inadequacy	5.3	13.3
12. IC1 Economic pressure	9.3	9.3
13. IB24 Hurt feelings	4.0	8.0
14. IA2a Pressure from partner	6.7	4.0
15. IA2s Unrequited love	4.0	4.0
16. IA2j Death or illness of partner	4.0	4.0
17. IA2e Aggression to partner	2.7	4.0
18. IB10 Rumination	1.3	6.7
19. IC5 Aggression from environment	2.7	4.0
20. IB7 Loneliness	5.3	2.7
III I Alternate Themes	9.3	10.7
III G Autobiographical	9.3	1.3
III K Denial of a theme	5.3	10.7

* Difference is significant at or beyond .05 level.

E. Identification

	H	NH
1. Confusion of sex		
a. Misidentification immediate	25.3	30.7
b. Uncertain at first and then misidentify	14.7	18.7
c. Uncertain at first and then identify correctly	4.0	10.7
d. Cannot decide whether male or female	8.0	1.3

More than half of all subjects are confused over the sex of the central character.

	H	NH
2. Identification of gun		
a. Questions whether it is a gun	14.7	9.3
b. Identifies as something else, e.g., scissors, comb, knife, etc.	5.3	13.3
c. Questions whether something else	2.7	5.3
d. Can't decide what it is	4.0	5.3

F. Attention to Details

	H	NH
1. Concern with object	40.0	42.7
2. Comments about figure, e.g., hunchback, deformed, etc.	9.3	17.3
3. Specify locale as prison	13.3	4.0
4. Specify locale as mental hospital	4.0	5.3

Card 4

A. Emotional Tone Ratings

	−2	−1	0	+1	+2
H	2.7	84.0	10.7	1.3	0
NH	4.0	94.7	1.3	1.3	0

Preponderance of stories is moderately sad.

B. Outcome Ratings

	−2	−1	0	+1	+2	?
H	6.7	10.7	21.3	40.0	1.3	18.7
NH	1.3	18.7	16.0	53.3	1.3	9.3

Most frequent outcome is moderately happy.

C. Shift

	Happier	Sadder	No Shift
H	57.3	5.3	37.4
NH	69.3	1.3	29.4

More than half of all subjects shift to happier conclusions.

Only rare individuals shift to sadder conclusions.

D. Most Frequent Themes

			H	NH
1.	IA2a	Pressure from partner	44.0	50.7
2.	IA2b	Succorance from partner	22.7	22.7
3.	IA2m	Competition	20.0	12.0
4.	IA2f	Departure from partner	12.0	16.8
5.	IA2r	Seduction	9.3	10.7
6.	IA3e	Aggression to peer	9.3	9.3
7.	IB11	Occupational concern	10.7	8.0
8.	IA2m	Jealousy of partner	16.7	6.7
9.	IA2h	Illicit sex	4.0	12.0
10.	IB6	Moral struggle	4.0	12.0
11.	IC1	Economic pressure	8.0	8.0
12.	IB4	Behavior disorder	8.0	6.7
13.	IB7	Guilt—remorse	6.7	4.0
14.	IB8	Drunkenness	6.7	4.0
15.	IA2p	Decision between partners	4.0	4.0
16.	IB1	Aspiration	8.0	0.0
17.	IB24	Hurt feelings	5.3	1.3
18.	IC5	Aggression from environment	4.0	2.7
19.	IC6	War	0.0	4.0

E. Identification of Characters and Objects

1.	Woman in background seen as real, not picture	10.7	8.0
2.	Woman in foreground seen as negress	2.7	6.7
3.	Woman in foreground seen as nurse	0.0	9.3

F. Attention to Details

1.	Mention of picture in background	34.7	42.7

Card 5

A. Emotional Tone Ratings

	−2	−1	0	+1	+2
H	1.3	17.3	72.0	2.7	0.0
NH	8.0	48.0	33.3	6.7	4.0

Most frequent story is neutral to moderately sad.

The hospitalized subjects contribute significantly more neutral stories and the non-hospitalized subjects significantly more moderately sad ones.

B. Outcome Ratings

	−2	−1	0	+1	+2	?
H	1.3	6.7	65.3	17.3	0.0	9.3
NH	6.7	13.3	37.3	17.3	5.3	6.7

Most frequent outcome is neutral.

Significantly more hospitalized subjects contribute neutral outcomes.

C. Shift

	Happier	Sadder	No Shift
H	22.7	2.7	74.6
NH	36.0	8.0	56.0

The majority of subjects do not shift for this picture.

D. Most Frequent Themes

			H	NH
1.	IB3	Curiosity	42.7	41.3
2.	IA1a	Pressure from parent	33.3	40.0
3.	IA2h	Illicit sex	8.0	20.0
4.	IA1g	Parental concern	10.7	12.0
5.	IB9	Fear	10.7	12.0
6.	IA2a	Pressure from partner	10.7	4.0
7.	IB4	Behavior disorder	2.7	8.0
8.	IA2g	Concern for partner	4.0	2.7
9.	IB17	Loneliness	4.0	2.7

Card 6 BM

A. Emotional Tone Ratings

	−2	−1	0	+1	+2
H	13.3	77.3	9.3	0.0	0.0
NH	22.7	77.3	0.0	0.0	0.0

Preponderance of stories are sad.

Neutral stories rare and contributed only by hospitalized subjects.

B. Outcome Ratings

	−2	−1	0	+1	+2	?
H	2.7	14.7	21.3	37.3	2.7	21.3
NH	17.3	17.3	20.0	46.7	2.7	10.7

Most frequent outcome is moderately happy.

Significantly more non-hospitalized subjects contribute extremely sad outcomes than do hospitalized subjects.

C. Shift

	Happier	Sadder	No Shift
H	58.7	2.7	39.6
NH	58.7	5.3	36.0

More than half the subjects shift to happier conclusions.

D. Most Frequent Themes

			H	NH
1.	IA1a	Pressure from parents	37.3	46.7
2.	IA1f	Departure from parents	24.0	36.0
3.	IA1p	Marriage of child	16.0	21.3
4.	IA1b	Succorance from parents	22.7	13.3
5.	IC4	Aggression towards environment	13.3	21.3

6. IA1m	Filial obligation	9.3	18.7
7. IA10	Bad news to parent	10.7	13.3
8. IA11	Death or illness of parents	10.7	10.7
9. IA1n	Confession to parent	5.3	16.0
10. IA1k	Disappointment to parent	10.7	12.0
11. IA1j	Death or illness of child	6.7	9.3
12. IA1g	Parental concern	10.7	2.7
13. IB11	Occupational concern	5.3	5.3
14. IA1c	Nurturance to parent	5.3	4.0
15. IB7	Guilt—remorse	1.3	6.7
16. IC5	Aggression from environment	6.7	1.3
17. IB16	Religion	4.0	2.7
18. IC1	Economic pressure	2.7	4.0
19. IC6	War	2.7	4.0
III I	Alternate Themes	6.7	6.7
III G	Autobiographical	2.7	4.0

Card 7 BM

A. Emotional Tone Ratings

	—2	—1	0	+1	+2
H	0.0	40.0	50.7	9.3	0.0
NH	5.3	58.7	32.0	4.0	0.0

Most frequent story is neutral to moderately sad.

B. Outcome Ratings

	—2	—1	0	+1	+2	?
H	1.3	9.3	34.7	40.0	1.3	13.3
NH	9.3	13.3	22.7	45.3	5.3	4.0

Most frequent outcome is moderately happy.

C. Shift

	Happier	Sadder	No Shift
H	41.3	4.0	54.7
NH	48.0	8.0	46.0

About half of the subjects shift, the majority to happier conclusions.

D. Most Frequent Themes

		H	NH
1. IA1b	Succorance from parent	80.0	74.7
2. IA1a	Pressure from parents	21.3	28.0
3. IB11	Occupational concern	21.3	25.3
4. IB1	Aspiration	9.3	9.3
5. IC4	Aggression towards environment	4.0	8.0
6. IB8	Drunkenness	5.3	5.3
7. IB2	Inadequacy	2.7	8.0
8. IIA1f	Contentment with parent	6.7	2.7
9. IB10	Rumination	2.7	5.3
10. IA1q	Collusion with parents	2.7	4.0
11. IA2a	Pressure from partner	1.3	5.3

Card 8 BM

A. Emotional Tone Ratings

	—2	—1	0	+1	+2
H	2.7	58.7	13.3	21.3	1.3
NH	13.3	40.0	9.3	37.3	1.3

About half of all stories are moderately sad.

B. Outcome Ratings

	—2	—1	0	+1	+2	?
H	1.3	6.7	10.7	54.7	5.3	18.7
NH	8.0	9.3	10.7	46.7	10.7	14.7

About half of all outcomes are moderately happy.

C. Shift

	Happier	Sadder	No Shift
H	50.7	4.0	45.3
NH	34.7	6.7	57.6

About half of all subjects shift, almost all positively.

D. Most Frequent Themes

		H	NH
*1. IB1	Aspiration	26.7	53.3
2. IC6	War	24.0	29.3
3. IA1i	Death or illness of parent	25.3	18.7
4. IB12	Death or illness of central character	21.3	14.7
5. IC5	Aggression from environment	13.3	8.0
6. IB14	Reminiscence, sad	9.3	6.7
7. IA3i	Death or illness of friend	10.7	2.7
8. IB11	Occupational concern	6.7	5.3
9. IA4k	Death or illness of sibling	8.0	1.3
10. IB10	Rumination	0.0	9.3
11. IC4	Aggression towards environment	5.3	4.0
12. IA3e	Aggression to peer	6.7	1.3
III N	Confused	6.7	1.3

* Difference is significant at or beyond .05 level.

E. Identifications of Characters and Objects

1. Misidentify sex of figure in foreground	1.3	1.3
2. Misidentify sex of figure on table	5.3	2.7
3. Background is actual scene	34.7	20.0
4. Background and foreground are two different events separated in time	1.3	4.0
5. Reverses so that figure on table dreams of figure in foreground	4.0	5.3
6. Figure in foreground "does not belong"	6.7	1.3
7. Figure on Table is Father	20.0	17.3
8. Figure on Table is Mother	4.0	1.3
9. Figure on Table is self	12.0	18.7
10. Figure on Table is friend	2.7	0.0

F. Attention to Details

1. Mention presence of gun	33.3	24.0
2. Give location of wound	8.0	4.0

Card 9 BM

A. Emotional Tone Ratings

	—2	—1	0	+1	+2
H	1.3	26.7	65.3	6.7	0.0
NH	0.0	26.7	48.0	24.0	1.3

Most frequent story is neutral.

Non-hospitalized subjects contribute significantly more moderately happy stories.

B. Outcome Ratings

	—2	—1	0	+1	+2	?
H	2.7	21.3	77.3	9.3	0.0	6.7
NH	1.3	12.0	56.0	20.0	2.7	8.0

Most frequent outcome is neutral.

Of questionable outcomes in NH group, 90% are contributed by psychoneurotics.

C. Shift

	Happier	Sadder	No Shift
H	20.0	9.3	70.7
NH	22.7	12.0	65.3

Less than one third of subjects shift from emotional tone of story to outcome.

D. Most Frequent Themes

			H	NH
1.	IIB4	Retirement	76.0	68.0
2.	IB21	Exhaustion	14.7	14.7
3.	IB19	Vacillation	9.3	13.3
4.	IIB2	Tranquility	4.0	16.0
5.	IC1	Economic pressure	9.3	9.3
6.	IC6	War	6.7	9.3
7.	IIA3b	Congeniality with peers	4.0	8.0
8.	IC3	Generalized restriction	5.3	6.7
9.	IC2	Legal restriction	5.3	2.7
10.	IA3b	Succorance from peer	4.0	4.0
11.	IB8	Drunkenness	5.3	1.3
12.	IC5	Aggression from environment	9.3	2.7
13.	IIB3	Reminiscence, happy	2.7	4.0
III	I	Alternate Themes	5.3	5.3

E. Identifications of Characters

	H	NH
1. Laborers	45.3	54.7
2. Tramps	32.0	29.3
3. Soldiers	8.0	14.7
4. Negroes	12.0	6.7

F. Attention to Details

	H	NH
1. Specify number of men	12.0	14.7

Card 10

A. Emotional Tone Ratings

	—2	—1	0	+1	+2
H	8.0	25.3	4.0	37.3	24.0
NH	12.0	48.0	1.3	24.0	17.3

About half the subjects contribute happy stories and half the subjects, sad stories. Only rare individuals tell neutral stories.

Significantly more non-hospitalized subjects tell moderately sad stories.

B. Outcome Ratings

	—2	—1	0	+1	+2	?
H	5.3	1.3	16.0	32.0	20.0	21.3
NH	9.3	8.0	8.0	41.3	21.3	6.7

Most frequent outcome is moderately happy.

Significantly more hospitalized subjects contribute questionable outcomes.

C. Shift

	Happier	Sadder	No Shift
H	28.0	6.7	65.3
NH	42.7	13.3	44.0

About half the subjects shift, significantly more to happier conclusions.

D. Most Frequent Themes

			H	NH
1.	IIA2c	Contentment with partner	54.7	45.3
2.	IA2c	Nurturance to partner	12.0	20.0
3.	IA2f	Departure from partner	10.7	21.3
4.	IIA2d	Reunion with partner	17.3	8.0
5.	IC6	War	8.0	16.0
6.	IA1j	Death or illness of child	6.7	8.0
7.	A1f	Departure from parent	6.7	6.7
8.	IIB3	Reminiscence, happy	5.3	5.3
9.	IA2g	Concern for partner	5.3	4.0
10.	IC5	Aggression from environment	2.7	6.7
11.	IIA1d	Reunion with parent	5.3	4.0
12.	IA2g	Pursuit of partner	5.3	2.7
13.	IB1	Aspiration	2.7	4.0
14.	IB12	Illness or death of central character	2.7	4.0

E. Identification of Characters and Objects

1. Two males	5.3	1.3
2. Two females	0.0	2.7
3. Young couple	20.0	13.3
4. Middle aged couple	6.7	8.0
5. Old couple	16.0	20.0
6. Mother and Son	8.0	12.0

Card 11

A. Emotional Tone Ratings

	—2	—1	0	+1	+2
H	10.7	41.3	40.0	1.3	0.0
NH	10.7	60.0	13.3	10.7	0.0

Most frequent story is moderately sad.

Significantly more hospitalized subjects give neutral stories and significantly more non-hospitalized subjects give moderately happy ones.

B. Outcome Ratings

	−2	−1	0	+1	+2	?
H	10.7	2.7	38.7	26.7	2.7	12.0
NH	22.7	13.3	25.3	21.3	2.7	9.3

Most frequent outcome is neutral.

Significantly more non-hospitalized subjects give sad outcomes.

C. Shift

	Happier	Sadder	No Shift
H	38.7	9.3	52.0
NH	32.0	14.7	53.4

More than half the subjects do not shift. Of those that do significantly more shift to happier conclusions.

D. Most Frequent Themes

		H	NH
1. IC5	Aggression from impersonal source	41.3	41.3
2. IC7	Escape from peril	30.7	24.0
3. IA3e	Aggression from peer	18.7	24.0
4. IB3	Curiosity	8.0	16.0
5. IA3b	Succorance from peer	2.7	5.3
6. IB9	Fear	4.0	4.0
7. IB16	Religion	5.3	2.7
8. IB12	Illness or death of central character	4.0	2.7
9. IC6	War	4.0	2.7
*III C	Description	17.3	4.0
III D	Unreal	5.3	9.3
III A	Symbolic	6.7	8.0
III E	Fairy Tale	1.3	12.0

* Difference is significant at or beyond .05 level.

Card 12M

A. Emotional Tone Ratings

	−2	−1	0	+1	+2
H	25.3	45.3	24.0	5.3	0.0
NH	26.7	42.7	28.0	2.7	0.0

Most frequent story is moderately sad; stories only rarely happy.

B. Outcome Ratings

	−2	−1	0	+1	+2	?
H	13.3	5.3	25.3	37.3	2.7	20.0
NH	17.3	9.3	24.0	40.0	4.0	5.3

Most frequent outcome is happy.

Of the non-hospitalized subjects, no normal subject gives a questionable outcome.

C. Shift

	Happier	Sadder	No Shift
H	48.0	6.7	45.3
NH	54.7	8.0	38.3

More than half the subjects shift to happier conclusions.

C. Shift

		H	NH
1. IA3m	Hypnotism	40.0	40.0
2. IB16	Religion	25.3	17.3
*3. IB12	Illness or death of central character	34.7	6.7
4. IB4	Behavior disorder	17.3	16.0
5. IIB5	Retirement	22.7	10.7
*6. IA1j	Death or illness of child	6.7	24.0
7. IA3b	Succorance from peer	17.3	10.7
8. IA3d	Aggression from peer	13.3	9.3
*9. IB3	Curiosity	1.3	13.3
10. IC5	Aggression from impersonal source	2.7	8.0
11. IA1a	Pressure from parents	4.0	6.7
12. IA1b	Succorance from parents	1.3	5.3
13. IA3a	Pressure from peer	5.3	1.3
14. IC4	Aggression towards impersonal source	2.7	4.0
15. IIA3c	Exhibition	1.3	5.3
16. IIB3	Reminiscence, happy	0.0	6.7
III I	Alternate Themes	12.0	6.7

* Difference is significant at or beyond .05 level.

E. Identification of Characters and Objects

	H	NH
1. Misidentify sex of character on bed	13.3	12.0
2. Characters are same age (young)	12.0	8.0

Card 13MF

A. Emotional Tone Ratings

	−2	−1	0	+1	+2
H	58.7	34.7	6.7	0.0	0.0
NH	45.3	49.3	5.3	0.0	0.0

Preponderance of stories are sad. Rare individuals contribute a neutral story.

B. Outcome Ratings

	−2	−1	0	+1	+2	?
H	18.7	25.3	13.3	21.3	2.7	20.0
NH	26.7	14.7	14.7	29.3	2.7	12.0

Most frequent outcome is moderately happy; but not significantly more frequent than moderately or extremely sad outcomes, which when taken together are significantly more frequent than happy outcomes.

C. Shift

	Happier	Sadder	No Shift
H	30.7	2.7	66.6
NH	46.7	10.7	42.6

About half the subjects shift, significantly more to happier conclusions.

D. Most Frequent Themes

		H	NH
1. IA2j	Death or illness of partner	60.0	42.7
2. IB7	Guilt—remorse	45.3	54.7

*3.	IA2h	Illicit sex	20.0	42.7
4.	IA2c	Aggression towards partner	16.0	17.3
5.	IB13	Retribution	13.3	16.0
6.	IB8	Drunkenness	10.7	13.3
7.	IC1	Economic pressure	14.7	9.3
8.	IA2i	Illicit sex with violence	10.7	9.3
9.	IB2	Inadequacy	13.3	4.0
10.	IB5	Suicide	10.7	6.7
11.	IA2u	Childbirth	8.0	5.3
12.	IB4	Behavior disorder	6.7	6.7
13.	IA2m	Jealousy	4.0	8.0
14.	IA2o	Cuckold	6.7	4.0
15.	IA2a	Pressure from partner	4.0	5.3
16.	IB21	Exhaustion	4.0	4.0
17.	IA3b	Succorance from peer	6.7	0.0
18.	IB16	Religion	4.0	2.7
19.	IC5	Aggression from impersonal source	4.0	2.7
III I		Alternate Themes	17.3	13.3
III G		Autobiographical	6.7	6.7
III K		Denial of a theme	5.3	2.7

* Difference is significant at or beyond .05 level.

E. Identification of Characters

1. Father-Daughter, or Mother-Son	6.7	1.3
2. Sister-Brother	4.0	1.3

F. Attention to Details

1. Mention books on table	12.0	13.3
2. Mention nudity of woman	12.0	12.0

Card 14

A. Emotional Tone Ratings

	−2	−1	0	+1	+2
H	9.3	18.7	45.3	22.7	4.0
NH	2.7	33.3	33.3	20.0	9.3

Most frequent story is neutral to moderately sad—not significantly greater than moderately happy.

B. Outcome Rating

	−2	−1	0	+1	+2	?
H	2.7	5.3	38.7	25.3	6.7	22.7
NH	2.7	13.3	26.7	36.0	16.0	9.3

Most frequent outcome is neutral to moderately happy.

C. Shift

	Happier	Sadder	No Shift
H	24.0	6.7	69.3
NH	33.3	9.3	57.4

Less than half of the subjects shift. Of those that do shift significantly more shift positively.

D. Most Frequent Themes

			H	NH
1.	IB3	Curiosity	25.3	26.7
2.	IB1	Aspiration	22.7	22.7
3.	IIB3	Reminiscence, happy	10.7	12.0
4.	IIB8	Ordinary activity	14.7	4.0
5.	IB4	Behavior disorder	4.0	14.7
6.	IB10	Rumination	5.3	12.0
7.	IIC1	Favorable environment	6.7	10.7
8.	IIB2	Tranquility	4.0	10.7
9.	IB17	Loneliness	8.0	5.3
10.	IC3	Generalized restriction	4.0	6.7
11.	IB5	Suicide	6.7	2.7
12.	IB11	Occupational dissatisfaction	4.0	5.3
13.	IB16	Religion	4.0	4.0
14.	IC6	War	2.7	5.3
15.	IC5	Aggression from impersonal source	2.7	5.3
16.	IB19	Vacillation	4.0	2.7
17.	IC1	Economic pressure	0.0	6.7
III A		Symbolic	1.3	10.7
III I		Alternate Themes	6.7	1.3
III G		Autobiographical	5.3	2.7
III J		Comments about artistic merits of picture, etc.	6.7	0.0

E. Identification of Characters and Objects

1. Individual climbing in window	8.0	0.0

Card 15

A. Emotional Tone Rating

	−2	−1	0	+1	+2
H	53.5	37.3	9.3	0.0	0.0
NH	52.0	42.7	4.0	0.0	0.0

Most frequent story is extremely sad. Neutral stories are rare.

B. Outcome

	−2	−1	0	+1	+2	?
H	20.0	16.0	22.7	16.0	4.0	28.0
NH	45.3	9.3	17.3	5.3	4.0	10.7

Most frequent outcome is sad.

Hospitalized subjects have significantly more questionable outcomes.

C. Shift

	Happier	Sadder	No Shift
H	40.0	4.0	56.0
NH	22.7	10.7	66.6

Less than half the subjects shift.

Significantly more shift to happier than sadder conclusions.

D. Most Frequent Themes

			H	NH
1.	IA2j	Death or illness of partner	45.3	32.0
2.	IB16	Religion	45.3	29.3
3.	IA3j	Death or illness of peers	18.7	21.3
4.	IB17	Loneliness	14.7	10.7
5.	IB4	Behavior disorder	5.3	16.0
6.	IB3	Reminiscence, happy	10.7	4.0
7.	IB7	Guilt—remorse	6.7	8.0
8.	IB9	Fear	6.7	6.7

9. IA1j Death or illness of child	2.7	8.0
10. IC6 War	5.3	5.3
11. IA3j Belongingness	0.0	10.7
12. IA3e Aggression towards peer	4.0	6.7
13. IA1i Death of parent	5.3	2.7
14. IA3b Succorance from peer	5.3	2.7
15. IB10 Rumination	2.7	5.3
16. IA4k Death or illness of sibling	4.0	2.7
17. IB3 Curiosity	0.0	6.7
18. IC4 Aggression towards environment	2.7	4.0
III D Unreal	10.7	17.3
III I Alternate Themes	6.7	9.3
III J Comments about artistic merits of picture	9.3	2.7
III A Symbolic	5.3	5.3
III B Abstract	2.7	6.7
III N Confused	6.7	0.0

E. Identification of Characters and Objects

1. Misidentify sex of character	4.0	4.0

Card 16

A. Emotional Tone Ratings

	−2	−1	0	+1	+2
H	2.7	17.3	38.7	20.0	5.3
NH	8.0	25.3	25.3	24.0	12.0

Most frequent story is neutral.

Moderately happy and moderately sad stories are of equal frequency.

B. Outcome Ratings

	−2	−1	0	+1	+2	?
H	4.0	4.0	38.7	24.0	4.0	9.3
NH	8.0	8.0	30.7	25.3	20.0	2.7

Most frequent outcome is neutral.

Significantly more non-hospitalized subjects contribute extremely happy stories.

No normal subject has a questionable outcome.

C. Shift

	Happier	Sadder	No Shift
H	21.3	9.3	69.4
NH	28.0	10.7	61.3

About two thirds of the subjects do not shift. Of those that do, significantly more shift to happier conclusions.

D. Most Frequent Themes

	H	NH
1. IIC1 Favorable environment	12.0	12.0
2. IC6 War	8.0	14.7
3. IIA2c Contentment with partner	8.0	5.3
4. IIB2 Tranquility	5.3	8.0
5. IIB7 Ordinary activity	6.7	5.3
6. IIA3b Congeniality with peer	6.7	2.7
7. IIA2c Ordinary Activity	2.7	5.3
8. IA1a Pressure from parents	2.7	4.0
9. IA1f Departure from parents	2.7	4.0
10. IA3b Succorance from peer	4.0	2.7
11. IB2 Inadequacy	5.3	1.3
12. IB10 Rumination	2.7	4.0
13. IC5 Aggression from environment	2.7	4.0
14. IC3 Generalized restriction	1.3	5.3
15. IIA1d Reunion with parents	2.7	4.0
III G Autobiographical	24.0	13.3
*III L Rejection	17.3	1.3
*III H Continuation of previous story	1.3	10.7
III C Description	6.7	1.3

* Difference is significant at or beyond .05 level.

Card 17 BM

A. Emotional Tone Rating

	−2	−1	0	+1	+2
H	2.7	24.0	38.7	30.7	4.0
NH	1.3	29.3	26.7	33.3	9.3

Most frequent story is neutral to moderately happy.

B. Outcome Rating

	−2	−1	0	+1	+2	?
H	8.0	8.0	30.7	33.3	4.0	16.0
NH	6.7	11.0	24.0	38.7	10.7	4.0

Most frequent outcome is moderately happy.

Significantly more hospitalized subjects contribute questionable outcomes.

C. Shift

	Happier	Sadder	No Shift
H	18.7	14.7	66.6
NH	21.3	16.0	63.7

About one third of the subjects shift and of these there is no significant difference in those shifting to happier or sadder conclusions.

D. Most Frequent Themes

	H	NH
1. IIBL Self esteem	14.7	30.7
2. IIA3e Exhibition	16.0	22.7
3. IIA3i Competition with peer	12.0	18.7
4. IC7 Escape from peril	17.3	17.3
5. IB3 Curiosity	16.0	8.0
6. IB1 Aspiration	14.7	5.3
7. IC2 Legal restriction	12.0	8.0
8. IIA3d Approbation from peers	6.7	12.0
9. IB18 Compensation	4.0	13.3
10. IIB2 Tranquility	4.0	8.0
11. IIB6 Resignation to lot	12.0	0.0
12. IC5 Aggression from impersonal source	6.7	4.0
13. IB2 Inadequacy	4.0	5.3
14. IB12 Illness or death of central character	1.3	8.0
15. IC6 War	2.7	5.3
16. IIB7 Ordinary activity	4.0	4.0
17. IB14 Vacillation	2.7	4.0
III I Alternate Themes	14.7	4.0
III G Autobiographical	5.3	5.3
III A Symbolic	2.7	4.0

E. Identification of Characters and Objects

	H	NH
1. Character is going up	46.7	40.0
2. Character is going down	22.7	29.3
3. Uncertain if character is going up or down	4.0	2.7
4. Character is going up and down	5.3	1.3

F. Attention to Details

1. Note muscles	26.7	24.0
2. Note nudity	13.3	8.0

Card 18 BM

A. Emotional Tone Ratings

	−2	−1	0	+1	+2
H	17.3	64.0	10.7	1.3	0.0
NH	24.0	74.7	5.3	2.7	0.0

Most frequent story is moderately sad.

Neutral and moderately happy stories are unusual.

B. Outcome Ratings

	−2	−1	0	+1	+2	?
H	21.3	9.3	26.7	25.3	0.0	17.3
NH	24.0	22.7	22.7	22.7	0.0	8.0

All categories of outcome are about equally frequent except questionable.

Of questionable outcomes in non-hospitalized group 90% are contributed by psychoneurotics.

C. Shift

	Happier	Sadder	No Shift
H	40.0	4.0	66.0
NH	36.0	10.7	63.3

More than half the subjects do not shift. Of those that do, significantly more shift to happier conclusions.

D. Most Frequent Themes

		H	NH
1. IB8	Drunkenness	37.3	49.3
2. IA3b	Succorance from peers	38.7	38.7
3. IA3a	Pressure from peers	21.3	26.7
4. IB4	Behavior disorder	14.7	25.3
*5. IC2	Legal restriction	28.0	12.0
6. IA3d	Aggression from peer	13.3	17.3
7. IB12	Illness or death of central character	17.3	13.3
8. IC4	Aggression towards impersonal object	13.3	10.7
9. IB7	Guilt—remorse	10.7	4.0
*10. IB13	Retribution	14.7	2.7
11. IC5	Aggression from impersonal source	4.0	9.3
12. IA3m	Hypnotism	4.0	5.3
13. IA1a	Pressure from parents	4.0	4.0
14. IA2a	Pressure from partner	4.0	4.0
15. IB2	Inadequacy	2.7	5.3
16. IB9	Fear	2.7	4.0
17. IB11	Occupational concern	4.0	2.7
III I	Alternate Themes	13.3	14.7
III A	Symbolic	5.3	6.7
III D	Unreal	4.0	2.7

* Difference is significant at or beyond .05 level.

E. Attention to Details

1. Dwell on number of hands	18.7	17.3

Card 19

A. Emotional Tone Ratings

	−2	−1	0	+1	+2
H	27	24.0	41.3	16.0	0.0
NH	9.3	33.3	14.7	33.3	2.7

Most frequent story is neutral to moderately sad.

Significantly more hospitalized subjects give neutral stories.

B. Outcome Rating

	−2	−1	0	+1	+2	?
H	2.7	2.7	32.0	21.3	0.0	25.3
NH	18.7	13.3	30.7	21.3	2.7	6.7

Most frequent outcome is neutral.

Hospitalized subjects give significantly more questionable outcomes and significantly less sad ones.

C. Shift

	Happier	Sadder	No Shift
H	17.3	6.7	76.0
NH	20.0	24.0	66.0

More than two thirds of subjects do not shift; and of these that do, there is no significant difference between the number giving sadder and happier conclusions.

D. Most Frequent Themes

		H	NH
1. IC5	Aggression from impersonal source	60.0	61.3
2. IIA1f	Contentment at home	20.0	26.7
3. IB19	Vacillation	8.0	8.0
4. IB9	Fear	2.7	10.7
5. IB12	Physical illness or death of central Character	0.0	6.7
6. IB4	Behavior disorder	0.0	8.0
7. IB16	Religion	4.0	2.7
8. IC6	War	1.3	5.3
III D	Unreal	13.3	25.3
*III C	Descriptive	14.7	2.7
*III L	Rejection	1.3	12.0
III G	Autobiographical	2.7	6.7
III I	Alternate themes	9.3	2.7
III B	Abstract	2.7	6.7

* Difference is significant at or beyond .05 level.

E. Identification of Characters and Objects

1. House in snow storm	44.0	54.7

2. Ship in storm	8.0	9.3
3. Igloo or quonset hut	4.0	4.0
4. House in flood	8.0	1.3
5. Ghosts and "spooks"	17.3	12.0

Card 20

A. Emotional Tone Ratings

	−2	−1	0	+1	+2
H	6.7	41.3	34.7	2.7	0.0
NH	13.3	74.7	9.3	2.7	0.0

Most frequent story is moderately sad.

Non-hospitalized subjects contribute significantly more sad stories and significantly less neutral ones than the hospitalized subjects.

B. Outcome Ratings

	−2	−1	0	+1	+2	?
H	9.3	10.7	36.0	24.0	0.0	17.3
NH	14.7	17.3	21.3	33.3	4.0	9.3

Most frequent outcome is neutral to moderately happy.

C. Shift

	Happier	Sadder	No Shift
H	33.3	8.0	58.7
NH	48.0	10.7	41.3

About half of all subjects shift, significantly more to happier conclusions.

D. Most Frequent Themes

			H	NH
1.	IB19	Vacillation	80.0	70.7
2.	IC1	Economic pressure	17.3	24.0
3.	IC5	Aggression from impersonal source	13.3	16.0
4.	IC4	Aggression to impersonal object	10.7	8.0
5.	IB8	Drunkenness	9.3	9.3
6.	IB11	Occupational concern	5.3	12.0
7.	IA3j	Belongingness	2.7	13.3
8.	IB17	Loneliness	6.7	10.7
9.	IIB7	Ordinary activity	9.3	6.7
10.	IA3b	Succorance from peer	6.7	8.0
11.	IB2	Inadequacy	5.3	9.3
12.	IB3	Curiosity	9.3	5.3
13.	IA2g	Unrequited love	8.0	5.3
14.	IC2	Legal restriction	9.3	4.0
15.	IC6	War	8.0	5.3
16.	IA2h	Illicit sex	2.7	9.3
17.	IB10	Rumination	4.0	8.0
18.	IB7	Guilt—remorse	5.3	4.0
	III I	Alternate themes	21.3	10.7
	III A	Symbolic	4.0	4.0
	III J	Comments about artistry	2.7	4.0
	III G	Autobiographical	5.3	1.3

E. Identification of Characters and Objects

1. Confusion of sex		
a. Misidentification immediate	6.7	14.7
b. Uncertain at first and then misidentify	0.0	5.3
c. Uncertain at first and then identify correctly	0.0	2.7
d. Unable to decide whether male or female	10.7	0.0

BIBLIOGRAPHY

1. Albee, G. W., & Hamlin, R. H. A preliminary rating scale of adjustment inferred from drawings. *J. clin. Psychol.,* 1949, **5**, 389-392.
2. Allport, G. W., Bruner, J. S., & Jandorf, E. N. Personality under social catastrophe: ninety life histories of the Nazi revolution. *Character & Pers.,* 1942, **10**, 1-22.
3. Anastasi, Anne, & Foley, J. P., Jr. An experimental study of the drawing behavior of adult psychotics in comparison with that of a normal control group. *J. exp. Psychol.,* 1944, **34**, 169-194.
4. Balken, E. R. A delineation of schizophrenic language and thought in a test of imagination. *J. Psychol.,* 1943, **16**, 239-271.
5. Balken, E. R., & Masserman, J. H. The language of phantasy: III. The language of the phantasies of patients with conversion hysteria, anxiety state and obsessive compulsive neuroses. *J. Psychol.,* 1940, **10**, 75-86.
6. Beck, S. J. *Rorschach's Test. Vol. I; Basic Processes.* New York: Grune & Stratton, 1944, 155-195.
7. Bellak, Leopold. *A guide to the interpretation of the Thematic Apperception Test.* New York: Psychological Corporation, 1947.
8. Bennett, G. Structural factors related to the substitute value of activities in normal and schizophrenic persons: I. A technique for the investigation of central areas of the personality. II. An experimental investigation of central areas of the personality. *Character & Pers.,* 1941, **10**, 42-50; 1942, **10**, 227-245.
9. Bettelheim, Bruno. Individual and mass behavior in extreme situations. *J. abnorm. soc. Psychol.,* 1943, **38**, 417-452.
10. Bondy, Curt. Problems of internment camps. *J. abnorm. soc. Psychol.,* 1943, **38**, 453-475.
11. Brown, J. F. *The psychodynamics of abnormal behavior.* New York: McGraw Hill, 1940, 3-14.
12. Bumke, O., & Kant, F. Cited in Jellinek, E. N. (ed.), *Alcohol addiction and chronic alcoholism.* New Haven: Yale University Press, 1942, pp. 98-99.
13. Cameron, Norman. *The psychology of the behavior disorders.* New York: Houghton Mifflin, 1947, 1-186.
14. Child, I. Note on Grant's new statistical criteria. *Psychol. Bull.,* 1946, **43**, 558-561.
15. Coleman, William. The Thematic Apperception Test: I. Effects of recent experience. II. Some quantitative observations. *J. clin. Psychol.,* 1947, **3**, 257-264.
16. Eron, L. D. Frequencies of themes and identifications in the stories of schizophrenic patients and non-hospitalized college students. *J. consult. Psychol.,* 1948, **12**, 387-395.
17. Eron, L. D., & Hake, D. T. Chapter 15-A, Psychometric approach to the evaluation of the Thematic Apperception Test. In Zubin, Joseph, & Young, K. M., *Manual of projective techniques.* Madison: College Typing Co., 1948.
18. Eron, L. D., Hake, D. T., & Callahan, Robert. The use of rating scales with the TAT. Unpublished manuscript.
19. Garfield, S. L., & Eron, L. D. Interpreting Mood and Activity in Thematic Apperception Test stories. *J. abnorm. soc. Psychol.,* 1948, **43**, 338-345.
20. Garrett, H. E. *Statistics in psychology and education.* New York: Longmans Green & Co., 1949, pp. 377-387.
21. Gesell, Arnold, & Amatruda, C. S. *Developmental diagnosis.* New York: Paul B. Hoeber, 1947.
22. Goldfarb, Wm. Psychological privation in infancy and subsequent adjustment. *Amer. J. Orthopsychiat.,* 1945, **15**, 247-255.
23. Greenson, Ralph R. Psychology of apathy. *Psychoanal. Quart.,* 1949, **18**, 290-302.
24. Grings, W. W. The verbal summator technique and abnormal mental states. *J. abnorm. soc. Psychol.,* 1942, **37**, 529-545.
25. Harrison, R. Studies in the use and validity of the Thematic Apperception Test with mentally disordered patients: II. A quantitative validity study. III. Validation by the method of "blind analysis." *Character & Pers.,* 1940, **9**, 122-138.
26. Harrison, Ross. The Thematic Apperception and Rorschach methods of personality investigation in clinical practice. *J. Psychol.,* 1943, **15**, 49-74.
27. Hertz, M. R. *Frequency tables to be used in scoring responses to the Rorschach inkblot test.* Cleveland: Western Reserve University, 1946, pp. 7-160.
28. Jacques, Elliott. The clinical use of the Thematic Apperception Test with soldiers. *J. abnorm. soc. Psychol.,* 1945, **40**, 363-375.
29. Kannenberg, K. N. A comparison of results obtained from the Thematic Apperception Test under two conditions of administration. *Amer. Psychologist,* 1948, **3**, 363 (abstract).
30. Klebanoff, S. G. Personality factors in alcoholism as indicated by the Thematic Apperception Test. *J. consult. Psychol.,* 1947, **11**, 49-54.

31. Kutash, S. B. Performance of psychopathic defective criminals on the Thematic Apperception Test. *J. crim. Psychopath.*, 1943, **5**, 319-340.
32. Leitch, Mary, & Schafer, Sarah. A study of the Thematic Apperception Tests of psychotic children. *Amer. J. Orthopsychiat.*, 1947, **17**, 337-342.
33. Lindquist, E. F. *Statistical analysis in educational research.* New York: Houghton-Mifflin Co., 1940, pp. 41-46.
34. McDowell, E. D. *Educational and emotional adjustment of stuttering children.* New York: Columbia Univ. Press, 1928.
35. Masserman, J. H., & Balken, E. R. The clinical application of phantasy studies. *J. Psychol.*, 1938, **6**, 81-88.
36. Morgan, C. D., & Murray, H. A. A method for investigating phantasies: The Thematic Apperception Test. *Arch. Neurol. Psychiat*, 1935, **34**, 289-306.
37. Murphy, Gardner. A comparison of manic-depressive and dementia praecox cases by the free association method. *Amer. J. Insanity*, 1921, **77**, 545-558.
38. Murphy, Gardner. Types of word association in dementia praecox, manic-depressives and normal persons. *Amer. J. Psychiat.*, 1923, **79**, 539-571.
39. Murray, H. A. Techniques for a systematic investigation of phantasy. *J. Psychol.*, 1937, **3**, 115-143.
40. Murray, H. A. et al. *Explorations in Personality.* New York: Oxford University Press, 1938, 530-545.
41. Murray, H. A. *Manual for the Thematic Apperception Test.* Cambridge: Harvard University Press, 1943.
42. Nirembersky, M. Psychological investigation of a group of internees at Belsen Camp. *J. ment. Sci.*, 1946, **92**, 60-74.
43. Potter, E. H. *An analysis of the effect of temporary situational factors upon Rorschach test results.* Unpublished Ph.D. dissertation, Yale University, 1948.
44. Rapaport, David, Gill, M. M., & Schafer, Roy. *Diagnostic psychological testing.* Chicago: Year Book Pubs., 1946, Vol. II, pp. 395-459.
45. Rautman, A. L., & Brower, E. War themes in children's stories, *J. Psychol.*, 1945, **19**, 191-202.
46. Rotter, J. B. Thematic Apperception Test: suggestions for administration and interpretation, *J. Personality*, 1946, **15**, 70-92.
47. Rotter, J. B. Studies in the use and validity of the Thematic Apperception Test with mentally disordered persons: I. Methods of analysis and clinical problems. *Character & Pers.*, 1940, **9**, 18-34.
48. Royal, R. E. Drawing characteristics of neurotic patients using a drawing-of-a-man-and-woman technique. *J. clin. Psychol.*, 1949, **5**, 392-396.
49. Sarason, S. B. Dreams and Thematic Apperception Test stories. *J. abnorm. soc. Psychol.*, 1944, **39**, 486-492.
50. Sarason, S. B. The use of the Thematic Apperception Test with mentally deficient children: I. Study of high grade girls. *Amer. J. ment. Def.*, 1943, **47**, 417-421. II. Study of high grade boys. *Amer. J. ment. Def.*, 1944, **48**, 169-173.
51. Sarason, S. B. *Psychological problems in mental deficiency.* New York: Harper Bros., 1950.
52. Stein, M. I. *The Thematic Apperception Test: an introductory manual.* Cambridge: Addison-Wesley Press, 1948.
53. White, R. W. *The abnormal personality.* New York: Ronald Press, 1948, pp. 102-174.
54. Wyatt, Frederick. Formal Aspects of the Thematic Apperception Test. *Psychol. Bull.*, 1942, **39**, 491 (abstract).
55. Yule, G. U., & Kendall, M. G. *An Introduction to the theory of statistics,* 11th ed. London: Charles Griffin & Co., 1937, pp. 534-535.

VOLUME 64
NUMBER 10

WHOLE NO. 316
1950

Psychological Monographs:
General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

Experimentally Induced Variations in Rorschach Performance

By
EDITH LORD
Mental Health Division
Arizona State Department of Health

Accepted for publication, April 23, 1950

Price $1.00

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION
1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

Copyright, 1950, by the
American Psychological Association

ACKNOWLEDGMENTS

Acknowledgment is made to the following persons for valuable suggestions and criticisms of this study: J. P. Guilford, William Grings, Bruno Klopfer, Louis P. Thorpe, Georgene Seward, S. M. Wesley, George R. Bach, Saul Rosenzweig, and Herbert S. Conrad.

CONTENTS

I. INTRODUCTION 1

II. THE PROBLEM 3

III. PROCEDURE 6

Administration 6

IV. RESULTS OF THE EXPERIMENT 8

Behavior on pre-tests 8

Variation in number of responses 8

Erlebnistyp 9

Statistical treatment of the data 11

Statistical findings 12

Review of results related to original hypotheses 24

V. DISCUSSION OF RESULTS 26

Experimental variables 26

Implications for further investigation 30

VI. SUMMARY AND CONCLUSIONS 32

APPENDIX 33

REFERENCES 34

I. INTRODUCTION

During the past decade many new personality tests have been devised, little-used old tests have been revised, popular tests have been polished, refined, and expanded. College catalogues throughout the United States have lengthened lists of courses offered in psychology to include instruction in the administration and interpretation of personality tests in general, of "projective techniques" in particular, and of the Rorschach Ink Blot Test specifically.

Literally hundreds of research reports and observational articles have been published on the Rorschach and its clinical use. The bulk of this literature has been concerned with Rorschach patterns as they relate to psychiatric syndromes, to social or cultural groups, to successful and unsuccessful students, workers, or military groups. The practical or empirical success of the instrument has been widely reported. The operational results of the Rorschach as a diagnostic instrument have been repeatedly acclaimed.

There is, however, a dearth of laboratory investigation of this important and widely-used clinical tool. In fact, there is even a question as to whether the test actually is a projective technique. Schachtel (31) offers an excellent critique of the projection hypothesis of the Rorschach test, and Bellak (3) states flatly that Rorschach's test is *not* a projective method. More clarification of these philosophical and theoretical areas underlying the Rorschach is definitely indicated.

Too, the paucity of statistical studies of reliability and validity, or of adequate norms, plus the existence of contradictory reports on these crucial points has given rise to a diapason of criticism. This cacophony has been particularly painful to the ears of the conscientious clinical psychologist who daily uses the Rorschach for diagnosis, for recommendations for therapy, for vocational or educational guidance and counseling, etc.

As Hsü (16) points out, there are two popular and disparate attitudes toward the Rorschach held by equally psychologically sophisticated persons. One maintains that the prime importance of the instrument is its revelation in symbolic terms of the personality of the individual, considered as a unique universe; the other stoutly maintains that this test, like all other psychological measuring devices, should meet at least minimum standards of statistical reliability and validity.

Hunt (17) criticizes the over-emphasis in clinical psychology on quantitative data rather than qualitative, and bases his hopes for the future progress of clinical testing on an increased attention to qualitative behavior of the person tested rather than on quantitative results of the test.

Qualitatively, does a subject, in fact, satisfy one of the basic assumptions underlying all projective techniques—that his concepts reveal his consistent way of organizing experience, as measured by the ideas and feelings he projects into "meaning-free" or ambiguous stimuli? Is a subject reacting only to the "unstructured" or "semi-structured" ink blots on the Rorschach test, or is he reacting to the total perceptual and social field, which includes the administrator of the test and the affective tone of the administrative situation? If these latter two variables are altered, will the subject's perception of the ink blots vary significantly? Can a subject conceal his way of organizing experience or simulate a pattern not his own? Can an administrator induce an altered Rorschach pattern by explicit or implicit suggestion?

There have been some very interesting explorations of these last two questions. Fosberg (11) studied attempts to fake Rorschach results, and also (12) examined protocols obtained under varied instructions. He demonstrated that psychologically sophisticated subjects, given a neutral or control Rorschach, could not conceal their basic test patterns on two subsequent tests in which they tried to give the best, then the worst, possible impression.

Schachtel (30) emphasizes the importance of a subject's personal definition of the test situation and its demands. He also strongly presents his personal conviction that the three most common subjective definitions of the Rorschach situation are as authoritarian, competitive, or resistance situations. These definitions, he maintains, influence Rorschach results.

Levine (22, 23) reports a study of a series of protocols produced by the same subject while experiencing different situations hypnotically induced. There were marked changes in the Rorschach records of the various situations, each taking on the coloring considered characteristic of a person experiencing the situation created by hpnotic suggestion.

Luchins (27) studied the influence on Rorschach responses of situational and attitudinal factors. Among his findings was the marked need for the development of methods which would prepare a subject for the Rorschach test. A significant factor in the Rorschach results he analyzed proved to be misunderstandings of directions and subjects' hazy concepts of what was expected of them in the testing situation.

Wilkins and Adams (33) studied results of Rorschachs obtained under hypnosis, under sodium amytal, and in the normal, waking state. They found that both the chemically and psychologically induced abnormal states contributed to increased productivity on the Rorschach among subjects described as "overly cautious" or "fearful."

An interesting variation in the sort of study conducted by Levine is that of Bergman, Graham, and Leavitt (4) who reported the Rorschach records of a hypnotically regressed patient, diagnosed as a conversion hysteric. They found that the protocols varied with the level of hypnotic regression but closely followed the expectancies growing out of other clinical data.

Cofer (9) studying the changes in responses given by a patient with organic involvement while under hyoscine, found that some of the cards are more sensitive than others to situational variations.

A recently published research project of the Army Air Forces (32) included an analysis of the influence of the examiner on the number of responses obtained on Rorschach records. Of 36 *t*-ratios between the means of nine examiners, 12 were significant at the 1 per cent level of confidence and three at the 5 per cent level. Clearly this one scoring symbol (R)[1] would seem to vary with the examiner.

Keeping this last finding in mind, and hypothesizing a continuum between the findings of Levine and Fosberg, one wonders whether the non-hypnotized subject ever experiences attitudes toward the administrator or mood reactions to the total administrative situation which are adequate to alter the determinants of his Rorschach record significantly. This, in brief, is the problem which led to the development of the present study.

---

[1] An explanation of the Rorschach scoring symbols used in this study is given in the Appendix.

II. THE PROBLEM

STATEMENT OF THE PROBLEM. This study of the influence of negative and positive rapport conditions on Rorschach performance is based on the hypothesis that the perception of ambiguous or "unstructured" stimuli is not influenced by the perceiver's affective reaction toward the total stimulus field, which includes the administrator—the person presenting the "unstructured" stimuli. If an experimental test proves this hypothesis false, implications are legion for the interpretation of test results and for the training of administrators.

Of importance is the possibility that this study—a test of the foregoing hypothesis—may reveal whether or not the administrator is a significant part of the stimulus field during the administration of projective tests.[1] The theory behind projective testing is that the subject structures stimuli presented to him in non-structured or semi-structured form. If a subject is projecting meanings, ideas, and feelings into ambiguous stimuli uninfluenced by the person presenting the stimuli or by the affective tone of the administrative situation, there should be no significant variation in his structuring of the stimuli if either the administrator or the affective tone of the administration is varied.

The problem of transference arises in all therapies, regardless of type, and the findings of this project might conceivably throw some needed light on the subject of transference-elements. If the requirements for positive transference could be determined or the situations in which it flourishes could be measured, the selection and training of therapists could be more clearly defined. It is thought that the present study may hold implications for this related problem.

Outside the realm of clinical psychology, this study may have implications for social psychology. Too, the manner in which inter-personal relationships affect perceptual processes is of great current interest to systematic psychology (14) as well as to the field of clinical psychology. While the present study does not answer the questions here posed, it does provide implications for future research in these areas.

Within the framework of the therapeutic situation, there may be at work individual variables which cause one subject to function more "normally" in a threatening environment, another to function more productively or adequately in an environment of heightened acceptance or permissiveness. In so far as the therapist's personality may structure these mood-situations, the results of this project could be revealing.

Designing an experiment to test whether Rorschach records will vary with experimental variation of administrator and tone of administration, one must ask himself, from a theoretical point of view, what variations might be expected. Remembering the experiments of Levine, one might hypothesize radical shifts in protocols from one administration to another if the administrations were adequately loaded with different affective tones. Remembering the experiments of

---

[1] Cognizance is taken here of the fact that one cannot amplify specific findings related to the Rorschach ink blots into generalizations concerning projective techniques as a category; however, it is thought that any light on the variability or stability of one projective technique may, at least, suggest areas of desirable investigation for all other clinical tools falling within the "projective" category.

Fosberg, wherein psychologically sophisticated subjects proved incapable of successfully simulating the Rorschach records of personality-types other than their own, one might predict no significant variation in total protocol, regardless of administrator or of type of administration.

Somewhere between these extremes, one may ask what sort of variation might be expected (a) if a group of subjects were given control Rorschach administrations, later were given positively loaded[2] Rorschach administrations, and still later negatively loaded administrations; (b) if a group of subjects were given Rorschachs by three different administrators at intervals of four to six weeks. Following are the hypotheses which seem to follow from the cited relevant studies, and which seem reasonable from the standpoint of personality dynamics and the Rorschach symbols or projective response-categories which are widely believed to be related to these dynamics.

1. There will be a variation in the number of responses elicited by the various administrators, regardless of affective variable or sequence of administration. This hypothesis is based on the A.A.F. study, cited above.

2. There will be no consistent variation in number of responses between the first, second, and third administrations; i.e., no increase or decrease in number of responses with successive administrations. However, the variation in number of responses will be related to a subject's initial performance. If a subject gives thirty or fewer responses on his first test, he will increase the number with subsequent testings; if he gives seventy or more on the first, he will decrease his responses on subsequent testing.[3]

3. Even among "normal" subjects, the alterations attendant upon variations of affective tone of administration will be largely dependent upon individual differences, the more stable personality records showing less change, the more labile records showing more change. It is predicted that approximately a third of the subjects' records will show no differences or only minimal changes regardless of administrator, order of administration, or experimentally varied affective tone of administration; a third will be extremely variable, fluctuating with administration order, administrator, and tone of administration. This stability or lability will be reflected in the constancy or shift of the M : sumC ratio.[4]

4. Regardless of the order of administration or the affective loading, there will be some variation in the records obtained by the three administrators which will be a function of the different personalities of the administrators. Since this factor is an unknown variable, it is not possible to predict the anticipated variations specifically. This hypothesis is a logical outgrowth of the known variation of the number of responses with examiner differences (32). Responses are scored as to location areas and determinants. An increase in responses insures an increase in one or another of these scoring categories. Furthermore, since number of responses has been shown to vary with examiner differences, it is not

---

[2] The word *loaded* is here used to mean *weighted* or *slanted*. See IV, C for complete exposition of this term as herein employed.

[3] Since thirty to seventy responses is usually considered the normal range, this hypothesis is based on an anticipation of the operation of central tendency.

[4] The particular choice of thirds in this hypothesis is based on verbal prediction made by Dr. Bruno Klopfer.

unreasonable to hypothesize that other scoring categories of the Rorschach may be equally sensitive to administrator differences.

5. Regardless of the administrator or the sequence of administration, there will be variations in the protocols produced under the three different administrative situations: neutral or control, positive affective loading, negative affective loading. Specific variations anticipated are the following:

a. Negatively loaded administration.

(1) Greater scattering of response determinants among the signs k, K, Fc, and C′, indicating a withdrawal from the painful situation (k or K), a depression of a situational sort (C′), and a sensitive approach to the total stimulus environment (Fc) as a way of handling anxiety provoked by the situation. These signs will tend to increase at the cost of the movement and color responses. Fm will appear with greater frequency as an indicator of tension in the face of a threatening situation.

(2) There will be an increase in white-space responses (S) as an indication of negativistic or oppositional tendencies.

(3) There will be an increase in F% and in FK + F + Fc/R as a reflection of an increased effort to master the situation through intellectual control or through this control supported by heightened sensitivity and introspection.

b. Positively loaded administration.

(1) There will be a greater emotional expression, or emotional relationship with the environment in this more permissive situation; this will be reflected in an increase in sumC.

(2) There will be an increase in movement responses, M and FM, indicating a wider play of creativity and a greater freedom of expression of "Id-drives."

III. PROCEDURE

An experimental design was constructed which aimed toward controlling as many variables as possible. The basic unit of the study was 36 males between the ages of 19 and 27. They were selected at random from a list of approximately 200 college sophomores enrolled in an introductory course in general psychology. Each subject was given three Rorschachs at four to six weeks' intervals. Each of the three Rorschachs was administered by a different person.

The three administrators, A, B, and C, were females, not different in any grossly apparent way; i.e., no one of the three was outstandingly fat or thin, ugly or beautiful. Administrators A and C were brunettes, B was a reddish-blonde. The age range among the three was twelve years, with B the youngest and C the oldest. Each administrator holds a degree in psychology and has had several years' experience in the use of the Rorschach according to the Klopfer technique.

The Rorschach tests were given in a systematic or rotated order so that twelve subjects received control tests first, twelve second, and twelve third. Twelve received negative tests and twelve received positive tests first. The second and third testing sessions were similarly varied.

No subject was tested twice by the same person. Twelve were tested first by Administrator A, second by B, and third by C. The other administrations were distributed similarly. Each administrator used exactly the same technique for the administration of the Rorschach Test, regardless of the affective loading of the pretest situation and regardless of whether the administration was the first, second, or third for the subject.

The negative and positive situations included two simple card-sorting tests administered prior to the Rorschach test for the purpose of setting the affective tone of the session. Identical directions, terminology, and techniques were used in both the negative and positive pre-tests; however, the tone of the administrations was differently slanted. Specific directions for the administration of the three sessions were followed explicitly and without variation by the three administrators. These directions were as follows:

ADMINISTRATION

Neutral. Standard Rorschach procedure will be used. The administrator will be courteous but business-like in manner. She must attempt to avoid either negative or positive affective loading of the situation. There will be no pre-tests nor any gathering of biographical data.

Negative administration. The administrator will assume the role of a harsh, rejecting, authoritarian figure. She must be deliberately unconcerned about the subject, not look at him while asking questions, preparing tests, or giving directions; never smile, give directions in a voice of dictatorial harshness, make every "Hm!" sound like a sneer.

Positive administration. The administrator will be personally warm, charming, appreciative in manner. She must look at subject with a smile while asking questions, preparing tests, or giving directions in an encouraging tone of voice, making every "Hm!" sound like a compliment for work well done.

The following administration will be used verbatim for *both* negative and positive administrations:

Are you Mr. Blank? Come with me! This room. Sit here.

What is your full name? You have an address? Freshman, Sophomore, Junior? Your birthdate?

Are you familiar with an ordinary deck of playing cards? (Place deck face down, with one black nine on the bottom of the pack.) When I say GO turn the cards, one at a time, and pull out all the black nines and the red tens. Put the black nines here and the red tens here (indicate areas with gestures). Work as fast as you can. Ready? Go! (Answer any questions only by repeating words in the foregoing instructions. Note any cheating, i.e., number of times subject turns two or more cards at a time; also note any errors in carrying out instructions or any requests for repetition of instructions.) Record time.

Hm! Some speed! (Leave nines and tens; shuffle remaining cards; replace, face down, before subject.)

Now, when I say GO, turn the cards, one at a time, and see how many face cards—Jacks, Queens, and Kings—you can pull out before I say STOP. Put the red face cards on the black nines and the black face cards on the red tens. Ready? Go! (Allow thirty seconds.)

STOP! (Count face cards and record number.) Hm! Not bad!

(The Rorschach Test will follow using the same instructions as for the neutral administration.)

As the experimental design reveals, each subject follows a unique pattern. Consequently, if a subject were lost at any stage of the experiment, a substitute would have to be started from the very beginning. In order to insure completion of the design, the entire pattern was started in duplicate; i.e., seventy-two subjects were given first Rorschachs. Despite this precaution, there were several cases wherein both of the pair were lost, and a third subject (with his parallel substitute) was started through the design. The principal cause for loss appeared to be the subject's reactions to the negatively administered session. Following this session, the subjects seemed more difficult to contact for testing appointments and frequently broke appointments or failed to appear as scheduled. Thirteen of the original seventy-two subjects were dropped following the negatively administered test because they had failed to keep from three to five appointments for the subsequent test. Two subjects were dropped for the same reason following positively administered sessions. Eight of the thirteen were given negatively loaded tests during the first session, five during the second. The losses, by examiner, were approximately equal, two losing four subjects each, the third losing five after negative administrations. The only reason for dropping these subjects, rather than persisting until an appointment was kept, was the time factor. If a period longer than six weeks elapsed between sessions, the subject was dropped and his parallel substituted in the pattern.

With this evidence in mind, one may speculate that the differences between negative results and positive or neutral results might have been greater if methodology had permitted retention of the negative records of the subjects who apparently rejected the entire project after experiencing the negative administration.

A few subjects were lost for other reasons. Three moved from the community before completing the three sessions; two were ill between sessions and unable to meet the requirement of a maximum time-lapse of six weeks between tests.

In every other case, testing of parallel partners was dropped as soon as the principal subject had completed his pattern and thereby precluded the need for a substitute.

As Rorschach records were accumulated, they were coded and mixed. The one hundred eight records were then scored without reference to subject, name of administrator, type or order of administration. Only main responses and determinants were used in the analyses of the records for two reasons: (1) incorporation of additional responses and determinants would unprofitably overcomplicate the study; (2) there are marked differences among exponents of the Rorschach method (e.g., Beck, Buhler, and Klopfer) as to the most desirable use to make of these additional determinants in scoring and in interpretation.

IV. RESULTS OF THE EXPERIMENT

In addition to the loss of subjects following negative administrations, there is other evidence that the negative situation was actually traumatic to the subjects. Although the very simple card-sorting pre-tests were designed merely to set the administrative tone of the administration for the Rorschach, behavior on the pre-tests suggests that in many cases the subjects were reacting differently by the time the pre-tests were introduced. The initial greeting, without a smile or a direct look, plus the manner and tone of voice in which the four personal-data questions were asked, apparently caused the subjects some concern.

A. Behavior on Pre-Tests

Not a single subject on the positively-loaded administrations misunderstood the pre-test directions or made an error in carrying them out. Twelve of the thirty-six subjects made from one to six errors on the negative administrations, with a total of twenty-five errors. Failure to understand directions was counted as an error if the subject requested repetition or clarification. If he demonstrated his confusion through incorrectly carrying out the directions, he was not corrected, but the error was noted.

During the Rorschach administration, the writer recorded every audible sigh or audible laugh. The laughs and sighs elicited by this one administrator give further qualitative evidence that the subjects were, in fact, reacting differently to the different affective loadings of the administrations: Eleven of the thirty-six subjects sighed aloud from one to seven times (total of thirty-two sighs) during the free-association period of the Rorschach, i.e., during the initial presentation of the plates. One subject sighed aloud during the positive, two during the neutral, and eight during the negative administrations. Seven subjects laughed aloud once or twice with a total of nine audible laughs; three laughed during the positive, two during the negative, and two during the neutral administrations. No effort was made to tabulate smiles and merely visible sighs.

One observes that audible sighs occur with greater frequency during a cold or rejecting situation than during a warm, permissive situation or a neutral situation, but that laughing aloud occurs with approximately equal frequency during the three different situations.

B. Variation in Number of Responses

The hypothesis concerning successive increase or decrease of number of responses as a function of the number given on the initial test held well for two cases which initially gave more than seventy responses but failed to hold consistently for twenty-six cases giving fewer than thirty responses at the outset.

There is an apparent bias in testing response limits exceeded by such an uneven number of subjects as exceeded our limits; i.e., two versus twenty-six. The explanation of this procedure rests in the earlier hypothesis based on erroneous original expectancy as to number of responses. Rorschach literature quite generally states that thirty responses can be considered the minimum expectancy with normal, intelligent, adult subjects, seventy responses the maximum expectancy. The failure of our subjects to conform to this expectancy is discussed later under the subheading, "Number of responses."

C. ERLEBNISTYP

Hermann Rorschach's theory of personality rests on the foundation of what he called *Erlebnistyp,* a word which has been variously translated but which is consistently rendered *Experience Type* in the English translation of Rorschach's original paper. He (29) makes reference to the problem, so labeled, throughout his *Psychodiagnostik,* and insists that an individual's protocol should be interpreted within the framework of his *Erlebnistyp,* that is, his introversive, extratensive, or ambiversive orientation, his "experience-balance" in responding to stimuli from within and/or stimuli from without. Rorschach found that *Erlebnistyp* was reflected in the ratio between movement and color responses (M : sumC).

The scoring symbol M denotes, according to Rorschach, "Form perceptions plus kinaesthetic factors"; (29, p. 25) and he adds that "The more kinaesthesis, the more stable the affectivity" (29, p. 26). "Color responses," he continues, "have proved to be the representative of the affectivity and the rule is, the more color in the test, the greater the emotional instability of the subject" (29, p. 76).

A person with more M than C in his record would be designated the M-type (introversive), described by Rorschach (29, p. 81) as having the following characteristics:

1. Predominance of personalized productivity.
2. Intensive rapport.
3. Stable affect and motility, awkwardness, insufficient adaptability to reality and insufficient extensive rapport.

A subject with more C than M would be designated the C-type, more extratensive than introversive. The general characteristics of this type, Rorschach (29, p. 83) lists as follows:

1. The urge to live in the world outside oneself.
2. Restless motility.
3. Unstable affective reactions.

Klopfer (21) has pointed out the necessity of taking into consideration other factors in the record before venturing more than a simple statement of the subject's "natural inclinations" as reflected in his M : sumC ratio. These additional considerations give information as to whether the individual follows these inclinations or is in a state of conflict over them.

For the purpose of the present study, we may confine ourselves to an analysis of the extent to which lability and stability of the M : sumC ratio exists within the framework of our experimentally varied testing situations.

Rorschach recognized the existence of temporary variations in number of M and C in a subject's record if the mood of the subject shifted between dejection and elation; however, he stated that despite the variation in absolute number, "the proportion between them (M and C) changes little or not at all" (29, p. 94).

In the 108 records produced by our thirty-six subjects, variation in absolute number ranged from 0 to 19 on the M side and from 0 to 13 on the sumC side. The proportion between the M and C was exceedingly shifty; and, in addition, there were numerous actual shifts of weight in the ratio from one side to the other. The number and direction of these variations in Experience Balance are summarized in Table I.

From Table I it can be seen that the greatest stability of M : sumC ratio exists between administrations I and II; even here, 31 per cent of the cases shifted from one balance to another; i.e., there was variation *within* approximately a third of the persons tested, as measured by an individual's production of more movement than color responses in one situation, more color than movement responses in another situation. A close second in rank order of stability is that between records obtained by administrators B and C, wherein 67 per cent maintained a constant balance.

The greatest number of shifts in balance occurred equally between Administrations II and III and between Administrators A and B. In both these categories, exactly half the cases maintained a constant balance and half shifted in one direction or another.

TABLE I

Number and Direction of Shifts in Balance on M: sumC Ratios

Situation	Frequency and Direction of Change M to C	C to M	A* to M	A to C	M to A	C to A	Stable	% Stable	% Unstable
I to II	1	2	0	3	4	1	25	69	31
I to III	5	4	2	3	3	0	19	53	47
II to III	6	6	1	4	1	0	18	50	50
A to B	7	2	1	2	4	2	18	50	50
A to C	8	2	3	2	1	1	19	53	47
B to C	2	4	2	4	0	0	24	67	33
0 to +	2	5	2	1	2	4	20	56	44
+ to −	4	5	3	2	2	0	20	56	44
0 to −	2	7	1	2	1	2	21	58	42
Totals	37	37	15	23	18	10	184	57	43

* The symbol A is used to denote ambi-equality of M and sumC.

In reading Table I, it must be remembered that a shift, from M to C for example, means that the subject actually showed more M than C under the first condition considered and more C than M under the second condition listed.

A first glance at the totals in the columns labeled "Per Cent Stable" and "Per Cent Unstable" might suggest that approximately half the subjects maintained a constant stability throughout the experiment regardless of sequence, administrator, or type of administration. Such a conclusion cannot be drawn from Table I, however, since stability therein is noted within, but not between, the variable situations.

An examination of the data reveals that only fourteen of the thirty-six cases maintained a constant balance on the M: sumC ratio throughout all variations in testing situations. Of these fourteen, ten held a balance weighted on the M side of the ratio, four showed consistent weighting on the sumC side, and none maintained ambi-equality of weighting. One may, therefore, conclude that 39 per cent of the subjects retained a stable Experience Type throughout the experiment, approximately three fourths of whom were introversively oriented, a quarter of whom were extratensively oriented.

These conclusions, then, would support the prediction stated earlier in this study that approximately a third of the subjects would produce stable records. However, on the basis of Rorschach's description of the M-type and C-type personalities, noted earlier in this discussion, one would have suggested that the fourteen subjects with stable records would *all* have been of the M-type. Table I shows us that A-type, M-type, and C-type personalities in one situation can become another personality type in another situation. The additional analysis of the data shows us that both M-type and C-type personalities are capable of maintaining their personality type in the face of varying situations, but that the odds are three to one that the subject who maintains such constancy will be an M-type personality.

D. Statistical Treatment of the Data

In order adequately to test the principal hypothesis of this experiment concerning the influence of negative and positive rapport conditions on Rorschach performance, and the numerous collateral hypotheses—the influence of successive administrations and different administrators, and the kind and amount of variation in the numerous scoring categories—one must look to statistical analysis.

To justify the greatest confidence in probabilities obtained from analysis of the data, *t*-ratios were computed for the numerous pairs of series offered by the data.[1] The particular *t*-ratio used was Fisher's formula as described by Lindquist (26, p. 58 ff.) under the title "The Significance of a Difference in the Means of Related Measures":

$$t = \frac{M_o - M_H}{\frac{\text{Sum } d^2}{n(n-1)}}$$

These *t*-ratios (a total of 243) and the interpretations of them are presented in the following pages.

In working the *t*-tests, data were fractionated in the following manner: All tests given during the first session were grouped, regardless of who gave them or of what kind of affective tone marked the administration; this procedure was followed for the second and third administrations as well. All tests given by Administrator A were grouped, regardless of whether they were administered first, second, or third in the series, regardless of whether the tone of administration was positive, negative, or neutral. A similar grouping was made for tests given by Examiners B and C. All tests given during a positively loaded test-situation were grouped without respect to who administered them or to the serial number of the situation. Identical groupings were made for tests administered in the negative and the neutral situations.

This combination of heterogeneous data, of course, introduces a special problem in the interpretation of the results of the *t*-tests. If variance is found due to administrations, this variance is operating when testing the differences between administrators and between types of administrations. Where significant variance exists only within one of the three major groups of variables, this problem loses some of its importance. If, however, variance is shown to exist as a function of the repetition of the test, this variance will operate to enlarge the sigma, hence enlarge the denominator, of the *t*-tests between other variables. The influence would produce spuriously low *t*-ratios among the other groups of variables.

---

[1] Cognizance is taken here of the criticism to which this study is subject as a result of the practice of treating scoring categories as unique parts of a whole without reference to the whole of which they are an integral part. The writer can think of no other method of experimental approach to the investigation of the Rorschach and the personality factors it is judged to measure except by analysis of the separate elements which contribute to the whole. Recognizing the imperfection of the approach, the plea is offered that it remains the best of all possible existing approaches.

This weakness of the *t*-test, which inevitably exists when data are fractionated and combined as in the present study, is not a serious drawback if one bears in mind that *t*-ratios approximating zero (say, under 1.000) throughout one whole set of data strongly indicate no real differences; if one remembers, also, that any significant or very significant *t*-ratio in one of the three groups of variables within a set will contribute to spuriously low *t*-ratios among the other groups, and that, therefore, a marginal *t*-ratio would be significant if the variance due to other factors were partialed out.

One other element of the statistical treatment demands discussion: Computation of *t*-ratios has been made on mean frequencies of scoring categories regardless of whether these scores are expressed as percentages of the total responses or as absolute numbers. While the percentages are not seriously affected by the variation in total number of responses, the absolute numbers are. For example, if the positive administration produces more responses than the negative, there will not necessarily be any alteration in the per cent of whole responses; there will, however, necessarily be an increase on the positive administration in the frequency of certain of the response determinants. On a statistical level, one is eager to discover *which* determinants increase; on an interpretative level, one attempts to explain *why* one determinant increases and another does not, whether the increase is a concomitant of increased responses or not.

If there were a significant difference in number of responses from the first to the second to the third administration, any differences in tabulated scoring symbols would have to be held suspect as an artifact or the frequencies would have to be reduced to ratios. However, no such differences in R were found to exist.

E. Statistical Findings

Number of responses. The greatest constancy in mean number of responses proved to be that between administrations. It is interesting to note here that Rorschach literature is consistent in reporting that the normal number of responses is around thirty with normal ranges reported from fifteen to seventy-five.

Hermann Rorschach: Normal subjects generally give from 15 to 30 responses, rarely less than 15, often more than 30 (29, p. 21).

Klopfer and Kelley: The range of responses found most frequently in all large-scale investigations of adults seems to be between twenty and forty (21, pp. 208-209).

Brussel and Hitch: Normal is around 34 . . . . Normal range is from 25 to 75 (6, p. 3).

Bochner and Halpern: The average (number of responses) seems to fall between 20 and 50 (5, p. 71).

Charlotte Buhler gives a minus weighting of 3 for fewer than twenty-five responses (7, p. 11).

As an aside, one wonders if these published means and ranges are not a bit high. Only on the positive administration and under C's administration do our subjects reach or exceed means of thirty responses.

Further in this vein, it should be noted that the previously cited study conducted by the Army Air Forces (32) shows the average number of responses for servicemen undergoing classification testing to be only twenty. Perhaps this particular Army test situation stimulated an inhibition of responses. Perhaps, too, the special population of the present study experienced a special kind of inhibition of response.

Perhaps, on the other hand, the figure thirty is not the true mean for the population as a whole, is not the appropriate cut-off point for a normal number of responses. Possibly separate norm-groups should be established for various segments of the population.

Although there is an increase in the mean number of responses on both the negatively and positively loaded administrations over the neutral administration, with the positive eliciting more than the negative, the variability within each type of adminisration is large enough to keep the differences below the critical cut-off point for statistical significance higher than the 10 per cent level of confidence which exists between the positive and the neutral administrations.

Thre is a very significant difference, however, between the number of responses elicited by Administrators A and C. Whether giving a first, second or third administration, whether being negative, positive, or neutral, Administrator C obtained, on the average, a sufficiently larger number of responses virtually to rule out chance as a cause of the difference (1 per cent level of confidence).

Average time per response. Very significant differences were found in the average response time between the first test and both the second and the third tests. These differences can be explained rather easily in terms of familiarity with the task. On being first confronted with the Rorschach plates, the subjects uniformly spent more time formulating their concepts than they spent on subsequent reassociation with the stimuli. The reduction in response-time is uniform in direction from beginning to end of the series, despite the lack of a significant difference between the second and third administrations.

There is a consistent tendency toward a difference in response times between Administrator C and the other two administrators; in fact, the difference between A and C is significant at the 5 per cent level when rounded to two decimal places. Apparently the subjects, on the average, lingered longer over their responses when giving them to Administrator C.

Crude control. F% is considered to be one of the key determinants in a Rorschach protocol. It is the scoring symbol used to indicate the percentage of responses determined solely by the shape, contour, or outline of the blot. Briefly, it is interpreted as a measure of the control exercised by the subject. Buhler and Lefever (8) designate it as a measure of rational functioning. Excessive F% (50 per cent or above) is considered to be an indication of undue constriction, or an accompaniment of depression or anxiety.

While the means of the various measures of our subjects stay well under the 50 per cent upper limit for normalcy, we may speculate that significant increases in F% mean increased depression, anxiety, or constriction. No such significant increases are found as a function of the number or the type of the administration; however, it is interesting to observe that the lowest mean is found for the positive administration and the highest for the negative.

Again we find significant differences among administrators, A differing at the 1 per cent level with C and at the 5 per cent level with B. There is no difference between B and C.

Refined control and its components. The FK + F + Fc per cent is another measure of control; however, it is distinguished from F% alone by the quality of the control it reflects. Klopfer (21) has called the latter "crude control" and the former "refined control." Separately considered, the determinant FK (vista concept) is usually interpreted as reflecting introspection or insight; Fc (texture concept), as reflecting tact or sensitivity.

Although there is not a critically significant difference in the amount of refined control displayed in the various types of administration, the means show a consistent increase from the lowest mean on the positive administration to the highest mean on the negative administration, paralleling the tendency on F% or crude control.

A significant difference between the first and the third administration suggests several alternative possibilities. The mere repetition of the Rorschach probably brought about a degree of loss of spontaneity in responding to the stimuli. This was replaced successively more frequently by increased refined control of reactions to the stimuli.

While there is practically no difference between Administrators A and B in the average percentage of responses with this combination of determinants, Administrator C differs at the 10 per cent confidence level from A and differs significantly from B.

Further light may be thrown on the factors influencing these differences in the combination-score by inspecting the components that contribute to it. F% has already been analyzed. The raw data show a low frequency of FK and Fc responses. The responses, however, occur with adequate frequency and with sufficient approach to normality of distribution to permit application of the *t*-ratio statistic.

A *t*-ratio for Fc between Administrators B and C was not computed since the difference between means of .05 is so small as to insure a *t*-ratio of approximately zero. Despite the low frequency and small differences in means between A and the other two administrators, the variability was sufficiently small to produce a significant difference between A and C, and a very significant difference between A and B. The latter two administrators consistently elicited more Fc responses than did A. This observation would tend to rule out Fc as a contributing factor to the significant difference between B and C on the FK + F + Fc percentage. Considering, too, the insignificant *t*-ratio between B and C on F%, one might suppose that despite very low frequency and apparent scatter, the FK responses are the greatest single contributing factor to the FK + F + Fc per cent difference between B and C; however, this convenient supposition is not supported by the data, as will be subsequently shown.

It is interesting to note that F% (crude control) was sufficiently higher for A than for B or C to produce significant and very significant differences respectively; whereas, Fc (tact or sensitivity) was sufficiently *lower* for A than for B and C to produce very significant and significant differences respectively. It is not surprising, then, that when these two determinants, F and Fc, are summed with a third determinant FK the differences between A and both B and C disappear but the difference between B and C is exposed.

The scoring category FK is used for concepts incorporating vista, distance, perspective. There are no significant differences in FK on the successive administrations, although the means reflect a consistent decrease in the use of this determinant. Nevertheless, no inferences can be made from such data.

The lack of significant differences among administrators also permits no interpretation, inference, or conclusion. We may only say that our analysis of FK offered no clarification of the questions growing out of the findings for FK + F + Fc per cent. This latter category, then, appears to be a unique whole, not paralleling any one of its three components in its stability or variation under the varied situations.

The affective loading of the administrations had a significant degree of influence on the scoring category FK. Both the negative and the positive administrations show more mean FK than does the control administration. The difference between the neutral and the positive administrations is significant at the 5 per cent level of confidence.

The symbols K and k. Responses classified as K are interpreted as a sign of "free-floating" anxiety or of insecurity. Douglas (10) found the use of this type of response serving "as conversational material until something further can be found."

K responses remain quite stable through the successive administrations. There is a noticeable difference, however, in the amount of K elicited by the different administrators, B and C differing at the 10 per cent level of confidence.

Like K, responses classified as k are interpreted as reflecting feelings of anxiety, insecurity, or inadequacy; however, the two categories are differentiated by the control factors implicit in a k response, absent in a K response.

The low frequency of the determinant k is reflected in low means for the various measures. Neither the variation of administrator nor the type of administration affected this scoring symbol to a degree worth noticing. There is a consistent, though insignificant, reduction in the appearance of this determinant on the successive tests, each administration eliciting somewhat fewer such responses than the preceding.

The C′ determinant. Our thirty-six subjects produced, on the average, about one and a half achromatic color responses (C′) per record without variation from one administration to the next. There was variation in the C′ responses with type of administration. The increase from an average of one C′ on the control Rorschachs to an average of two on the positively loaded Rorschachs boosts the confidence level between these variables to 10 per cent. Administrator A elicited the lowest frequency of C′ responses, differing from both B and C at the 5 per cent confidence level.

Movement responses. There are no statistically significant differences higher than the 10 per cent level of confidence in the number of Fm, inanimate movement, responses produced under any of the three variations of the testing situation. The constancy in number from one administration to another is so great as to preclude the necessity of computing *t*-ratios.

Mean differences, however, exist as a function of the variation in type of administration, the *t*-ratio between the control and the positive administrations exceeding the 10 per cent level of confidence that a real, not chance, difference exists.

The animal movement responses, FM, vary little or not at all on the successive administrations, and they fail to vary significantly with variation in type of administration. However, there are differences at the 5 per cent level of confidence between Administrators A and C, and at the 1 per cent level between B and C, Administrator C eliciting the larger number of responses in both comparisons.

The frequency of responses classified as M, human movement, varied sufficiently to reach the 5 per cent level of confidence between one pair of variables and the 10 per cent level between four other pairs.

Human movement is consistently interpreted as a sign of inner adjustment or equilibrium, a capacity for the absorption of emotional stimuli whether originating from within or without.

An excess of M on the positive administration over the neutral created a difference significant at the 5 per cent level of confidence, and the positive administration shows a difference from the negative at the 10 per cent level.

Administrator C provoked the highest mean frequency of M responses, differing from both A and B at the 10 per cent level of confidence. Further, the mean number of M on the third administration was sufficiently below the number on the second to give a *t*-ratio significant at the 10 per cent level.

Emotionality. The C-type personality was discussed earlier in this paper, and the relationship of sumC to M was defined. Not forgetting the importance of this inter-relationship, let us look at the meaning of color responses per se in Rorschach records. Color is considered to be the determinant of emotionality. Lack of color is considered an indication of constriction; excessive use of color is interpreted as emotional lability or overreaction to external emotional stimuli.

The use of color varied little on successive administrations. There was an infinitesimal difference between the subjects' sum of color responses on the neutral and the negative administrations; in fact, no significant differences occurred among any of the types of administration.

The really important differences in the use of color responses occurred with the variation of examiners. Administrator A elicited significantly fewer color responses than C and very significantly fewer than B.

Content categories. Normally, the sum of animal and animal-detail concepts constitutes approximately 50 per cent of the content of Rorschach concepts. Our subjects linger close to this norm. An increase in A% is usually interpreted as evidence of stereotypy of thinking or a repression of intellectual activity; a decrease in A%, contrarily, is widely interpreted as less confinement to the obvious, a broader range of thinking and/or interests.

There are completely unimportant differences in A% on successive administrations of the test. There is no appreciable difference in A% from one examiner to another; although, again, the differences between Administrator A and the other two examiners show a stronger tendency toward significance than does the difference between B and C, that between A and C closely approaching the 10 per cent level of confidence.

The percentage of A% responses on the affectively loaded administrations differs significantly, with the negative administration producing the percentage higher than that on the positive administration. Also, the neutrally conducted tests deviate from the positive with a difference at the 10 per cent level of confidence.

There is agreement among users of the Rorschach that the number of animal responses should and do exceed the number of animal-detail responses in a normal, healthy record. Therefore, it was considered advisable to break down the A% scoring symbol into its component parts in an effort to examine the circumstances under which both animal responses and animal-detail responses may vary or remain stable.

Examination reveals no significant differences in the number of whole animal responses elicited in the various situations. Even the means do not show a difference worth comment. Certainly this analysis of the number of whole animal responses does not help appreciably to clarify the differences obtained in the A%. A look at the means and critical ratios for animal details is indicated; however, one must not overlook the fact that absolute number frequencies introduce different considerations from those under consideration when one is speaking of percentages.

In all measures, the number of A responses exceeds the number of Ad responses, indicating that our subjects on the average maintain the healthier balance regardless of variation in testing situations. With both symbols there are infinitesimal differences in frequencies from one testing situation to the next. Among administrators, C elicits the highest frequency for both A and Ad responses. Among the types of administrations, the control Rorschachs show the lowest mean for both scoring categories.

Differences at the 5 per cent level of confidence exist between Administrator A and both B and C, with A gleaning significantly fewer Ad responses than B or C.

As with other scoring categories, the number of A and Ad responses by themselves are considered to have little meaning. Whereas, an increase in Ad may indicate a tendency toward a critical attitude, this interpretation of excessive Ad in a record is made only if the sum of animal-detail and human-detail responses exceeds the sum of whole animal and whole human responses. This observation calls for an analysis of the number of human and human-detail responses.

Several measurable differences in number of whole human responses exist. The frequencies are exceedingly stable from the first through the third administration. While Administrator C provoked a larger mean number of human responses than did either A or B, only the difference with A is great enough to produce a *t*-ratio significant at the 10 per cent level of confidence.

The means of the types of administration show an increase in human concepts on both the negative and positive administrations over the number on the control administration, with the difference between the neutral and positive rising to the 10 per cent level of confidence.

Again, the frequency of whole versus part concepts of human figures is a more important consideration than the mere number of either.

A comparison of means reveals that the records of our subjects, regardless of variables, maintain a healthy preponderance of H over Hd responses.

Administrator C evoked a greater number of Hd responses than either B or A, and the mean of C's administrations differs significantly from that of A.

Popular responses. The means of the popular responses elicited by the various examiners are identical. There is a slight but steady increase in mean P from the first to the second to the third administration, which may easily be explained in terms of increased familiarity with the material through mere repetition.

The difference between the means of the negative and positive administrations reaches the 10 per cent level of confidence, the permissive situation stimulating the higher number.

Use of ground as figure. Neither repetition of administration nor affective loading of administration produced any difference in the number of white-space, S, responses on the Rorschach records. Neither are Administrators A and B essentially different in the amount of S they provoked on the protocols. Administrator C, however, differed from both A and B in eliciting a higher frequency of white-space responses, and differed from A at the 5 per cent level of confidence.

An earlier discussion of additional responses explained why they were not incorporated into the analysis of results on this study. An exception should be made, however, of the white-space scoring category. Among normal or near-normal subjects, white-space responses occur most frequently concomitant with reaction to inked areas of the plates. Arbitrary scoring methodology requires that, in such mixed concepts, the inked area be scored as the main response location and the white space be scored as an additional response location. Consequently, many clear-cut reactions to white space are buried among the additional responses. Better to understand the actual amount of and variability in the use of white space, it was decided to sum the absolute number of main and additional S responses and analyze the distribution of this total reaction to ground.

The total use of white spaces by our subjects varies significantly with type of administration and among administrators. The smallest amount of total S + (s) was elicited on the neutral or control administrations, the most on the positive administrations. While the actual mean difference between neutral and negative administrations is small, the direction of change is so consistent, the variability within the columns so small, that the difference reaches the 5 per cent level of confidence. The higher mean difference between neutral and positive administrations is accompanied by sufficiently increased variability to hold the difference down to a 10 per cent level of confidence.

Among administrators, there exist real differences in the amount of sum-S produced, regardless of number or type of administration. A evoked the least, and C the most, use of white space. The difference is significant at the 5 per cent level of confidence. C also exceeded B in elicitation of total S responses with a difference significant at the 10 per cent level. The most significant difference, however, is between administrators A and B, with a *t*-ratio at the 1 per cent level of confidence.

Percentage of responses to all-color cards. Color, as previously stated, is interpreted as a correlate of emotionality. Since cards VIII, IX, and X on the Rorschach test are composed exclusively of bright-color inks, they are construed to have particular significance in reflecting emotional components of a subject's personality. The normal expectancy of responses on these all-color cards is approximately 40 per cent of the total responses.

Our subjects consistently approximate normal expectancy, with little variation from one administrative situation to another. The largest difference occurs between the neutral and negative administrations, where the *t*-ratio rises to a 10 per cent level of confidence.

Apparently the over-rejecting situations brought about a measurable decrease in response to external emotional stimulation, creating a dampening effect on the subjects.

Manner of approach. The number of whole responses in a Rorschach record provides two separate clues for interpretative analysis: (1) a clue as to the percentage of responses which are based on incorporation of the total blot-stimulus in the concept-formation; (2) a clue as to the ratio between whole responses and movement responses, i.e., the W : M ratio.

The first of these clues will be explored later upon the presentation of means and *t*-ratios for W%. The second requires the reintroduction of the means for the scoring symbol M.

There is a steady decrease in the mean number of whole responses with each repetition of the test, the difference between number of W responses on administrations I and III being significant at the 5 per cent level of confidence.

If the W : M ratio much exceeds 2 : 1, the interpretation is usually made that the subject is over-extending himself; i.e., he is striving beyond his means of achievement. It is not considered a healthy sign. Let us compare our W : M mean ratios for any evidence of this imbalance.

One ratio that was obtained on the second administration achieves the ideal relationship of 2 : 1. With one exception, all other ratios show heavier W weighting; however, the ratio of 3 : 1 is not sufficiently in excess of the ideal to merit interpretation of a significant imbalance. The ratio on the neutral administration of 4 : 1, however, is twice the expectancy ratio. On the basis of Rorschach literature, one would have predicted a 2 : 1 ratio on this control administration, regardless of what other predictions attended the other administrations.

The really crucial interpretative factor related to whole responses is the scoring symbol W%—the percentage of total responses which use all of the inked surface of the cards within the framework of a single concept. "In a normal record," say Brussel and Hitch, "about 25 per cent of whole answers are expected" (6). The Individual Record Blank for scoring Rorschach protocols, developed by Klopfer and Davidson, and in wide use, shows a normal range of W% extending from 20 to 40. Every mean in the present study exceeds not only the figure given as a normal mean but also the figure given as the upper limit in the range of normal expectancy for W%. One might assume that the stated norms are in error; however, it seems less arrogant to consider the possibility that the subjects in the present study deviate from the population as a whole somewhat.

Examination supports the earlier finding of reduction of whole responses as a mere function of repetition of the test. When working with percentages rather than absolute numbers, this reduction in tendency is more dramatically revealed. While the decrease in W% is insignificant from the first to the second administrations, the decrease is significant at the 1 per cent level between the third and both the first and the second administrations. C's protocols show the lowest mean percentage of whole responses and differ at the 5 per cent level of confidence with A's records, which show the highest mean W% among administrators.

While none of the differences in means of varied types of administration are statistically significant, it is interesting to observe that the negative administration shows the highest percentage of whole responses and the positive situation the lowest percentage.

As W% decreases, it is mathematically certain that other location sub-divisions will increase. The nature of these increases can be explored only by analyzing the frequencies in the data of large details, small details, and tiny or unusual details as they are employed by the subjects in concept-formations.

One clear-cut difference exists between Administrators A and C. Administrator A elicited the highest percentage of W% responses and the lowest percentage of D% responses; in both scoring categories, A and C differ at the 5 per cent level of confidence.

No other variation in testing situation creates differences of statistical significance. Nevertheless, consistent tendencies are revealed by the means of the various types of administrations. The negative administration provoked the lowest mean D% and the positive administration the highest. The *t*-ratio between them barely misses reaching the 10 per cent level of confidence.

The location scoring symbol D is used when subjects form a concept using a large, insular part of the blot, frequently perceived as a separate entity by normal subjects. Between 45 per cent and 55 per cent of total responses are expected to be D responses.

Not until the third administration did the subjects show any measurable increase in the use of small usual details in the blots. The production of concepts based on these small areas remained constant from the first to the second test but differed between the second and the third to a degree significant at the 10 per cent level of confidence.

The average intelligent subject is expected to have from 5 per cent to 15 per cent of his responses in the d% category. The third administration, then, most nearly provoked d% responses up to a minimum of expectancy for our intelligent subjects. Decreased d% is not generally interpreted as having any particular significance, and increased d% is ignored unless it exceeds 15 per cent of the total responses.

Although there is a lack of significant differences between the various administrators, Administrator C has the highest mean d% within the administrator group, just missing 10 per cent *t*-values with both A and B.

Dd + S% is a sort of waste-basket category for all responses based on blot areas not defined as W, D, or d. Such responses may include tiny details, inside or edge details, unusual combinations of large or small blot areas, and the use of ground as figure. White-space responses were analyzed earlier by absolute number because of their special qualitative meaning in Rorschach interpretation. Their tabulation in combination with unusual details is a necessary part of the process of recording percentages of responses by location areas alone.

There is complete agreement among Rorschach authorities that absence of Dd + S% is normal. There is, however, some difference of opinion as to how high this percentage may go before the upper limit of normalcy is exceeded. The range of estimates is from 2 per cent to 10 per cent, with the consensus approaching the larger figure. In any event, the records of normal persons rarely exceed 10 per cent in this category. On all variations of the testing situation, the subjects in this study remain under this 10 per cent maximum; however, the ceiling is approached in the means of third administration and of the positively administered tests. None of the differences between means are statistically significant.

While Dd + S per cent in excess of 10 per cent is interpreted as an abnormal sign, even an excess of 5 per cent in the category is, by some authorities, considered to be evidence of a pedantic approach to the test (and, by extension, to the environment and to external stimuli, in general). All nine of our means exceed 5 per cent. The foregoing interpretation is more readily made if the higher percentage of Dd + S responses occurs in conjunction with high W%, a fact which we have already observed to exist with our group.

Summary of t-*ratios.* This completes the statistical analysis of the data on the numerous Rorschach scoring categories. The somewhat lengthy analysis of means and *t*-ratios by Rorschach scoring categories seemed the most clear-cut way of presenting the data of this study. However, that procedure inevitably produces some confusion as to how many statistically significant *t*-ratios were found and where, how frequently Administrator A or administration + had the highest means and on which types of response. These data are presented in summary form in Tables II to IV.

TABLE II

Rank Order of Means, According to Type of Response

Type of Response	Administration			Administrator			Type of Administration		
	I	II	III	A	B	C	+	—	o
R	2	1	3	3	2	1	1	2	3
T	1	2	3	3	2	1	1	3	2
F%	3	2	1	1	2	3	3	1	2
FK	1	2	3	2	3	1	1	2	3
K	1	2	3	2	3	1	1	2	3
Fk	1	2	3	2	3	1	1	3	2
#F	3	1	2	1	3	2	2	1	3
FC	1	3	2	3	2	1	1	2	3
CF	1	2	3	3	1.5	1.5	1	3	2
C	1	3	2	1	2	3	1	3	2
FK+F+Fc/R	3	2	1	2	1	3	3	1	2
M	2	1	3	2	3	1	1	2	3
sumC	1	3	2	3	1	2	1	3	2
FM	2	1	3	2	3	1	1	3	2
Fm	2.5	2.5	1	3	2	1	1	2	3
Fc	3	2	1	3	2	1	1	3	2
C′	1	2	3	3	2	1	1	2	3
A%	2	3	1	1	2	3	3	1	2
P	3	2	1	2.5	2.5	1	1	3	2
H	2	1	3	2	3	1	1	2	3
A	2	1	3	2	3	1	1	2	3
Hd	3	1	2	3	2	1	1	2	3
Ad	3	2	1	3	2	1	2	1	3
CR%	2	3	1	3	1	2	2	3	1
W%	1	2	3	1	2	3	3	1	2
#W	1	2	3	3	2	1	1	2	3
D%	3	2	1	3	2	1	1	3	2
d%	2	3	1	3	2	1	2	3	1
Dd+S%	3	2	1	3	2	1	1	2	3
S	1	3	2	3	2	1	1	2	3
S+(S)	1	3	2	3	2	1	1	2	3

Table II serves to show the administrative circumstances under which the means of each of the thirty-one Rorschach response categories, herein analyzed, held the highest, middle, or lowest numerical position within its group.

Table III shows the frequency with which each variable, within each of the administrative groups, achieved the highest, middle, and lowest means. It will be noted that the positive administration and Administrator C reveal almost identical frequency patterns. No other pairs show this marked parallel tendency.

As Table IV reveals, there were, at the 1 per cent level of confidence, a total of ten differences, over four times as many significant *t*-ratios as one would expect by random sampling; four of these were a function of repetition of the test, six a function of examiner differences. Administrator A differed very significantly from

B on three of the scoring categories and from C on two of the measures. B and C differed only once at this level of significance.

At the 5 per cent level of confidence, one would expect twelve significant differences by random samplings; whereas, we obtained twenty-one additional differences at this level alone. Combined with the ten *t*-values at the 1 per cent level, our total is thirty-one *t*-ratios equal to or above the value required for confidence at the 5 per cent level. This figure is almost three times chance expectancy. Successive administrations accounted for two of the significant differences, and type of administration for four. The fifteen differences among examiners point up the greater frequency between A and C than between either A and B or C and B. Three of the administrator differences were between A and B, one between B and C, but eleven existed between A and C. Combining differences at the 1 per cent and 5 per cent levels, A differs six times with B and thirteen times with C, while B and C differ a total of twice.

TABLE III

Frequency of Mean Position, According to Administrative Variable

Administration	Highest Mean	Middle Mean	Lowest Mean
I	13	$8\frac{1}{2}$	$9\frac{1}{2}$
II	7	$15\frac{1}{2}$	$8\frac{1}{2}$
III	11	7	13
A	5	$8\frac{1}{2}$	$17\frac{1}{2}$
B	$3\frac{1}{2}$	19	$8\frac{1}{2}$
C	$22\frac{1}{2}$	$3\frac{1}{2}$	5
+	23	4	4
−	6	14	11
o	2	13	16

Differences (5 per cent level) resulting from variation in the emotional tone of the testing situation occurred twice between the neutral and negative administrations, once between the positive and neutral, and once between the positive and negative administrations.

Clearly, the largest number of significant and very significant differences at the 1 per cent and 5 per cent levels were the result of examiner differences; and, within our group of examiners, A was the greatest contributing factor to frequency of differences. The negative and neutral administrations appear with equal frequency (three times each) as contributing factors to significant differences in type of administrations. The positive administration is a factor in two of the four differences.

Table IV summarizes the seventeen *t*-ratios at the 10 per cent level of confidence. This figure brings to forty-eight the total number of *t*-values reaching or exceeding the 10 per cent confidence level, a total exactly twice as large as the twenty-four one would expect by random sampling.

At this level of significance, 10 per cent, the variations with type of administration predominated, with nine differences tending toward significance. Two differences were a function of repetition of the test, and six occurred with change of administrators. Administrator C was a party

TABLE IV

Number of Meaningful *t*-Ratios by Variables

Confidence Level	I II	II III	I III	A B	A C	B C	+ −	+ o	− o	Cumulative Total
1%	1	0	3	3	2	1	0	0	0	10
5%	0	1	1	3	11	1	1	1	2	31
10%	0	2	0	0	3	3	2	6	1	48

to all of these variations, differing three times with A and three times with B. Of the nine differences occurring with variation in the tone of the administration, the positively loaded situation differed six times with the neutral administration and twice with the negatively loaded situation. The negative and positive administrations showed differences at this level of confidence only twice, the negative and neutral only once. Summarizing, the positive administration was a factor eight times, the neutral seven times, and the negative three times.

Summary of results by scoring categories. A few of the Rorschach scoring categories proved exceedingly stable, showing no variation from test to retest, from examiner to examiner, from positive to negative to neutral administration. These stable categories are the following: Dd + S per cent, Fk, and number of A responses.

The majority of the Rorschach scoring categories showed measurable variability as the testing situation varied. The degree of variation, together with the administrative variable which produced the variation in each shifting category, was as follows:

a. R: There was one difference at the 1 per cent level of confidence between administrators and one difference at the 10 per cent level with variation in types of affective loading of the administrative situation.

b. T: There were two differences at the 1 per cent level in average time per response on the successive administrations and one difference at the 5 per cent level between administrations.

c. M: This scoring symbol proved to be quite variable, with one difference at the 5 per cent level and another at the 10 per cent level as a function of the varied affective loading of the testing situation; two differences occurred at the 10 per cent level with variation in administrators, and one difference at the 10 per cent level grew out of mere repetition of the test.

d. FM: Neither successive administrations nor variation in affect in the testing situation influenced FM. Change of administrators, however, produced one difference at the 1 per cent level, another at the 5 per cent level.

e. Fm: This scoring symbol was resistant to change with repetition of the test or with change of examiners. It varied only with the variation in type of administration, and then only at the 10 per cent level of confidence.

f. K: Administrator differences produced one difference in K at the 10 per cent level of confidence. No other variables in the testing situations measurably influenced this determinant.

g. FK: This response determinant also remained stable through successive tests and despite variation in examiners, but it showed a 5 per cent level of significant difference with variation in affective loading of the testing situation.

h. Fc: Examiner difference again is evident with this determinant, there being both a 1 per cent and a 5 per cent difference among administrators. Neither retesting nor affective loading affected the mean number of Fc produced.

i. C′: This scoring symbol varied once at the 5 per cent level with repetition of the test, twice at the 5 per cent level with change in administrators, and once at the 10 per cent level with variation in tone of administration.

j. sumC: The composite score sumC remained stable from one test to the next and despite variation in affective loading of the testing situation; however, the mean scores varied significantly twice with change of administrators, once at the 1 per cent level, once at the 5 per cent level.

k. F%: This factor also resisted change with number and type of administration; however, it, too, varied with the examiner once at the 1 per cent level, again at the 5 per cent level.

l. A%: The difference in the type of administration produced two differences in A%, one at the 5 per cent level, the other at the 10 per cent level. Neither administrators nor successive testing influenced the means significantly.

m. P: The number of popular responses remained somewhat constant throughout the experiment. One difference, at the 10 per cent level, occurred with variation in the affective loading of the testing situation.

n. FK + F + Fc / R: The affective loading of the administration did not influence this composite scoring factor; however, both successive testing and change in administrators brought about one difference each at the 5 per cent level of confidence. Examiner differences accounted for an additional difference at the 10 per cent level.

o. H: The number of human-figure responses varied at the 10 per cent level once with change in administrators and once with variation in the type of administration.

p. Hd: Human-detail responses were unaffected by retesting or by affective loading of the test situation; however, there was one difference at the 5 per cent level with change in administrators.

q. Ad: Animal-detail responses were also impervious to the emotional tone of the administration and to successive testing, but they twice showed a difference at the 5 per cent level with variation in administrators.

r. CR%: The percentage of responses to the all-color cards remained somewhat stable, showing only a single difference at the 10 per cent level of confidence upon variation of the affective loading of the administration.

s. #W: The number of whole responses varied once at the 5 per cent level with successive testing but remained impervious to change with varied administrator or type of administration.

t. W%: W% remained unchanged with variation in emotional loading of the test situation, but showed two differences at the 1 per cent level with retesting and one difference at the 5 per cent level with variation in administrators.

u. D%: This location scoring symbol varied once at the 5 per cent level with change in examiners but otherwise remained quite unchanged.

v. d%: The percentage of small usual details remained stable except for one difference at the 10 per cent level with successive testing.

w. S: The number of white-space responses did not differ from one examination to the next but varied between examiners once at the 5 per cent level and differed once at the 10 per cent level with alteration of affective loading of the administration.

x. S + (s): The sum of main and additional white-space responses resisted change with successive administrations; however, the variation in affective loading produced one difference at the 5 per cent level. Change of administrators proved to be the most significant variable with one difference at the 1 per cent level, one at the 5 per cent level, and a third at the 10 per cent level of confidence.

F. Review of Results Related to Original Hypotheses

Let us briefly review the specific hypotheses and predictions made at the outset of this experiment and check our results against them.

1. The crucial hypothesis related to the stability of responses to the ink blots regardless of variation in environmental stimuli other than the ink blots; i.e., the person presenting the blots, the implicit attitude of the administrator, or the mere repetition in presentation of the blots. Of the twenty-three Rorschach elements subjected to statistical analysis, only three were relatively unaffected by the extra-test stimuli, by "situational" stimuli. Nineteen of the twenty-three elements entered into differences at least once at the 1 per cent, 5 per cent, or 10 per cent levels of confidence. These nineteen varying factors showed a total of forty-six differences attendant upon some variation in stimuli other than the test stimuli—the ink blots themselves. The hypothesis of stability does not hold.

2. It was predicted at the outset that the number of responses would vary with administrators, regardless of number or affective tone of the administration. The means of responses for the three administrators were, roughly, 24, 27, and 33, the difference between highest and lowest means being significant at the 1 per cent level of confidence. These findings support the prediction.

3. No consistent variation in the number of responses was anticipated from the first to the second to the third administration; and, in fact, there was none. Initial records of over seventy responses were expected to decrease with successive administrations, which they did. However, the prediction that initial records under thirty responses would show successive increase in number of responses was not supported by the data.

4. Stability as a function of individual differences reflected in M : sumC ratio proved, as predicted, to isolate approximately 30 per cent of the cases. Also fulfilled was the prediction that the more stable personality records, those of the M-type individuals, would show more situational stability than the records of the C-type persons. Thirty-nine per cent of the subjects retained a stable Experi-

ence Type throughout the experiment, approximately three-fourths of whom were M-type personalities, one-fourth C-type.

5. The hypothesis that there would be some variation in the records obtained by the three administrators, regardless of number or type of administration, was more than supported by the data. In fact, examiner differences proved the most pervasive influence in the experiment—perhaps the most dramatic finding of the study.

6. The predicted increase in frequency of k, K, c, and C′ with the negative administrations was in no particular realized. In fact, both Fk and Fc had the lowest mean frequency in the negative testing situation; the other symbols—K, FK, and C′—showed means on the negative administrations which were exceeded by the mean frequencies on the positive administrations. Apparently, as judged by these determinants, our thirty-six subjects found the positively loaded testing situation more challenging or threatening than the negatively loaded. This finding negates the hypothesis under discussion.

7. The hypothesized increase in white-space responses in the negative testing situation was somewhat supported, if one judges negative means against control means. Both main S and main-plus-additional S responses showed a mean increase in the negative administrations over the control administrations. However, the negative means of both measures were exceeded by the means on the positive administrations. Apparently the negative administrations did provoke more oppositional behavior than the control administrations; however, the positive administrations elicited even more of this type of response than either the negative or the neutral administrations, a finding which, in effect, negates the hypothesis of peak negativistic behavior on the negative administration.

8. Both crude and refined control, as measured by F% and FK + F + Fc per cent respectively, were expected to increase in the negative administrations. These predictions were supported by the data.

9. The anticipated increase in emotionality on the positive administration, as measured by sumC, was realized. There was a steady rise in mean sumC from the negative to the neutral to the positive administrations.

10. The increase in movement responses (M and FM) on the positive administrations occurred as predicted. Both scoring categories achieved their highest mean with the positive administration. The lowest mean frequency on FM occurred with the negative administration; on M, with the neutral administration.

V. DISCUSSION OF RESULTS

In the foregoing report of all of the significant findings for the experimental variables of this study—administrations, administrators, and types of administrations—there was little attempt to make qualitative evaluations or dynamic interpretations of the findings. Nevertheless, one is justified in attempting—in fact, is obligated to attempt—to relate these data to the dynamics of personality.

A. Experimental Variables

An effort will be made to summarize the effects produced upon the experiencing subjects by the different variables herein studied. Too, an attempt is made to understand the differences among the three examiners in terms of the different effects they so consistently produced in the test results. While these procedures have highly speculative components, every effort is made to hold the speculations close to the actual data, to translate into dynamic terms the actual findings of this experiment.

Influence of repetition of the test. The mere repetition of the Rorschach was the least important variable in the entire study, accounting for the smallest total number of important *t*-values, only eight in all. The categories showing change with successive administrations are, for the most part, such as can be quite easily accounted for. One is not surprised to find the average time per response consistently decreasing with repetition, this change accounting for two administrative *t*-values. On being successively confronted with the same stimuli, subjects needed no time to adjust to the stimuli; they could employ mere recall as a substitute for perception, apperception, and conceptualization. Furthermore, there is the strong possibility that the subjects, on once learning the total demand or expectancy offered by the test-challenge, simply functioned in line with our cultural speed-value, meeting the total requirements of the situation with as little expenditure of time as could be managed.

We have elsewhere indicated that our subjects frequently showed moderately compulsive tendencies. As a defense mechanism, the compulsive tendencies would be expected to be most operative when the subjects felt most challenged or threatened. One influence of repetition of the test on this dynamic pattern would be reduction of threat or challenge and consequent reduction in display of the compulsive defense. Three of our eight administration *t*-values grow out of reduced use of the whole blot in concept formation on successive tests. This phenomenon would seem to need no further comment.

Repetition of the Rorschach induced a successive and consistent increase in the use of "refined control." Again, one does not have to delve into deep dynamics of personality to understand this finding. Our intelligent subjects merely demonstrated their ability to improve their mastery over a challenging situation with each successive association with the test-task.

The remaining measurable influence of repetition on the test results is concerned with M, human movement responses. On taking the test for the second time, our subjects manifested greater equilibrium, more inner adjustment, a higher level of imaginative thinking. This is in line with the previously noted increase in security, the display of refined control over the situation which attended successive administrations. The third administration of the same test, however, failed to stimulate our subjects to display this sign of the superior, well-adjusted person. There is a measurable decrease in human movement responses on the third administration. One may speculate that the subjects were inhibited on the first administration, were less inhibited and more secure on the next association with the blots, hence free to demonstrate optimal capacity to absorb and creatively utilize stimuli, but were perhaps just bored with the third presentation of the same blots, bored to the point of not bothering to behave optimally.

The suggestion that boredom attended the third testing situation is somewhat supported by the spontaneous comments of many of the subjects during the third session: "These again?" "I've already done this twice." "Same old thing—a butterfly," etc., etc.

Effect of Examiner A. Among administrators, A elicited responses which reveal the following functioning of the subjects, as indicated by rank order of means: Fewest number of responses, shortest average time per response, least degree of emotional adjustment or social adjustment (FC), least amount of emotional lability (CF), smallest measure of emotionality (sumC), least expression of sensitivity (Fc), least tendency to use all possible environmental stimuli (C′), least stimulation of critical or perfectionist attitudes toward concepts (Ad and Hd), lowest frequencies in all location areas except W% wherein A's subjects show the highest mean.

The signs on which A's mean frequencies were highest in rank order suggest the following reactions to the effect of this administrator on the subjects: Greatest amount of crude control (F%), most frequent manifestation of unhealthy, uncontrolled direct emotional interaction with the environment (C), strongest tendency toward compulsive behavior (W%).

The pattern which seems to form, as one reviews these summarized effects of Administrator A on the subject's Rorschach records, somewhat resembles what one might expect of a person confronted with a threatening, frustrating situation. If this total effect may be considered a mirror of the administrator's personality, then Examiner A would be described as a cold, forbidding, frustrating, threatening figure, these personality components permeating the test-situation regardless of the deliberate and varied role-playing attempted.

There is no way of checking the validity of this thumb-nail sketch of A's personality components as reflected in her effect on Rorschach records. The writer, however, did ask two psychologically sophisticated persons, acquainted with all three examiners, to give subjective descriptions of each. Both described A as the coldest, most inflexible, and most solid of the examiners. One added the adjective "masculine" and the phrase "castrating type of female."

Effect of Examiner B. Administrator B's results fell in the middle position on rank order of means with the majority of measures analyzed in this study. The relative effect of her personality on the administrations is revealed in part, perhaps, by the very paucity of extreme positions achieved under her administrations. Following are the measurable effects: smallest amount of introspection (FK), lowest indication of maturity or creative imagination (M), and least evidence of bouyancy or free expression of primitive drives (FM).

The subjects, under B's administrations, achieved rank order of 1 on only two significant means of categories: highest total degree of emotionality (sumC) and most refined control (FK + F + Fc/R).

The pattern reflected in the mirror of effects suggests that B may possess personality traits that are emotionally exciting to the subjects; at the same time, this excitement apparently is well controlled and does not provoke undue anxiety or tension, neither does it stimulate intellectual activity.

It will be remembered that Administrator B is the youngest of the three examiners. She was subjectively described as the most feminine of the group, the "softest," most nearly the "ideal protecting mother-figure." One person described her as essentially "seductive."

Effect of Examiner C. Administrator C stimulated in subjects the least amount of either crude or refined control (F% and FK + F + Fc/R), the lowest measure of uncontrolled emotionality (C), and the least indication of "whole compulsion" (W%).

The means of the significant measures showing highest rank order under C's administrations are the following: Most responses with shortest average time per response, most introspection and unfixed anxiety (FK and K); best social and emotional adjustment (FC), most evidence of creative imagination (M) and of bouyancy (FM), highest indication of sensitivity (Fc), freest use of all possible stimuli (C′), most evidence of ease in relating to people (H), greatest evidence of critical or perfectionist attitudes toward productions (Ad and Hd), greatest use of all blot location areas except W% wherein C's subjects show the lowest mean.

The effect of Examiner C on the results of the Rorschach are more complex than the effects of the other examiners. One may speculate that C provided more of an intellectual than an emotional stimulation to the subjects. There would seem to be evidence of challenge with attendant anxiety, but there is also evidence of easy rapport, social adjustment, and less need for control devices in the situation with C.

C is the eldest of the three examiners. She was subjectively described as the most flexible of the three, as more feminine than A but less than B. One person described her as exuberant, "bubbling," the other as "very sympathetic."

Before leaving these speculations on the effects of the various examiners, it should be pointed out that Administrator A quite consistently approximates the pattern of the negative administrations with sixteen agreements out of a possible thirty-one. Administrator B most nearly approximates the pattern of the neutral administrations with fourteen agreements out of thirty-one. Administrator C almost duplicates the pattern of the positive administration with twenty-six agreements out of the possible thirty-one.

One may speculate that the basic personalities of the examiners, which permeated all testing situations regardless of the varied affective roles they played, are in fact related: A = negative; B = neutral; C = positive. Or one might infer that Administrators A and B, for example, in deliberately structuring the positive administrations actually created a situation somewhat like that which was consistently and unintentionally created by C in all administrations, that A and C, in attempting to be neutral succeeded only in resembling B's essential personality, and that B and C in structuring negative situations succeeded merely in resembling A's basic pattern. The former interpretation seems more likely. Whichever interpretation appears more reasonable, the facts and figures serve further to underline the importance of examiner differences.

Effect of positive administrations. This highly permissive situation was deliberately structured to give subjects the feeling of acceptance; the atmosphere was one of approval from beginning to end. This procedure brought about significant changes in the Rorschach behavior of the subjects, as has been previously noted. Let us summarize the effect of these positive administrations:

More responses were elicited, suggesting greater productivity in the more permissive situation. These responses tended to incorporate all possible stimuli presented. There was a greater display of intellectual activity, of creative imagination, plus evidence of increased thinking of a popular or communal type. Stereotypy of thinking was markedly low during the positive administrations. There was evidence of increased resistance against feelings of personal inadequacy plus a rise in tension-tolerance and in self-questioning, or introspection. Finally, there was marked evidence of a greater ease in relating to other human beings.

Effect of negative administrations. The negative test-situations were non-permissive, rejecting; an atmosphere of disapproval prevailed throughout. The effect of this environmental tone on the Rorschach records is reflected in a definite increase in stereotypy of thinking and a decrease in thinking that is in harmony with community thinking. There is a low incidence of intellectual activity of the imaginative or creative sort. Too, there is evidence of withdrawal from emotional stimuli, as measured by a decrease in the number of responses to the all-color cards. There is a marked rise in self-questioning, or introspection, plus evidence of resistance against feelings of personal inadequacy.

Effect of neutral administrations. The chief value in the experimental design of the neutral set of Rorschachs is that of a control factor. Since these tests were administered in the standard manner, without affectively loaded pre-tests, they serve somewhat as a measure of the reactions to the Rorschach which one might expect to find under ordinary clinical situations. The results on the neutral Rorschachs, however, cannot be construed to have the same meaning as is ordinarily attributed to the results of control groups in psychological experimentation. A third of our subjects had had one affectively loaded Rorschach before experiencing the neutral situation; two thirds had had two affectively loaded experiences with the test before receiving the neutral administration. It would be folly to assume that all of these neutral administrations provided identical test-situations for our subjects, that there was no carry-over from previous experiences with the administrative tone in which the blots were presented. Nevertheless, our neutral administrations do offer a modicum of evidence as to how subjects react to the test when standard testing conditions prevail.

The neutral situation elicited fewer responses than the positive, and these responses showed higher stereotypy of content than occurred on the positive administrations. There was less interest in utilizing all environmental stimuli, diminished ease in inter-personal relationship, and diminution of intellectual creativity on the neutral—as compared with the positive—administrations. Lower tension-tolerance and less resistance to feelings of inadequacy were also apparent in the neutral, over the positive, administrations. A lower resistance reaction was evident on the neutral administrations when compared with the negative situations. Less introspection was apparent on the neutral than on the positive administrations, and there was greater reaction to the emotional stimulation of all-color cards.

B. Implications for Further Investigation

This study does not supply the answer to a question which grows out of the findings: What kind of personality was possessed by each of the three administrators? Since administrator differences appeared as the greatest contributing factor to the variation in the Rorschach records, one would like to know what specific and measurable factors in an examiner's personality will bring about what specific and measurable factors in a subject's Rorschach protocol. One wonders if a Rorschach psychogram of the three examiners would resemble the composite mean profiles of the subjects in every way or in any way. These problems should be investigated.

The present study was designed in such a way that a subject could withdraw without suffering too great a penalty. Consequently, several of the individuals who reacted intensely to the negative administration withdrew from the project without finishing the three tests. Their records, of course, could not be included in this study. This entire experiment should be repeated with a group of subjects over whom the examiner can exercise greater control. The fact that differences were revealed as a function of the negative administration makes one wonder how much more significant these differences might have been—and what additional differences might have appeared—if the records of the intensely affected subjects could have been incorporated in the data.

The experimental design of the present study required that each subject not only experience three different affective loadings of the test situation but also experience three different examiners. Two separate investigations should be made as a supplement to the present study. It would be well to know whether the same kind and amount of fluctuation of scoring symbols and M : sumC balance would occur if a single examiner tested a group of subjects under varied affectively loaded situations or if several examiners successively tested a group of subjects under the same affectively loaded circumstances. This latter study should, in fact, be broken into at least two studies, one investigating the results of a series of positively loaded situations, and the other exploring the results of a series of negatively loaded administrations.

In order to minimize the personal equation and thereby to study in purer form the influence of negative and positive rapport factors on Rorschach performances, one could repeat this experiment with the following variations: neutral, positive, and negative administrations could be administered in group form, according to the design herein used, with directions or instructions given by a single examiner through the medium of phonographic recordings. This examiner would, of course, have to cut three records, each affectively differently loaded. While such a procedure could not effectively incorporate the particular card-sorting pre-tests used in this study as tone-setters for the administration, other equally simple devices could be created. In fact, the findings of the present study suggest that no such elaborate preamble is necessary to instill feelings of acceptance or rejection. A cold reception proved adequate to disturb our subjects to the point of misunderstandings and errors.

The significant findings for the group of male college sophomores justify a repetition of this entire experiment with other age groups and with female subjects. While our subjects were randomly selected, they were not selected from the population as a whole but from a rather small stratum of a very small and highly selected population—college sophomores. Our group was shown to deviate from population norms in several instances; e.g., in compulsive tendencies and pedantic approach. One may only speculate on the extent to which other findings in this experiment are a function of the select group from which the subjects were drawn. Would a similar study, with subjects representing a cross section of the population as a whole, show more or less variation, show the same or different kinds of instability?

All of these questions, and many more, must eventually be answered in the laboratory if we are to refine our use of the most popular and pragmatically valuable of the projective techniques: The Rorschach Ink Blot Test. This study is a small step in a direction which must, and certainly will, be taken by other students who realize that the practical success of of the Rorschach must be supported by laboratory verification. This area is rich in experimental problems sufficient to excite at least a generation of students.

VI. SUMMARY AND CONCLUSIONS

This investigation has attempted to explore the stability or variability of the numerous response-determinants of subjects' concepts on the Rorschach Ink Blot Test, to discover the extent to which mere repetition of the test, a change of administrators, or an alteration of the affective tone of the testing situation would bring about alterations in number and content of responses, in determinants, and in location areas.

Related studies are extremely few and their contradictory results are explainable as a function of their decidedly different experimental designs. A single study, that of the Army Air Forces, is closely related to the present experiment in its analysis of variation in number of responses as a function of examiner-differences.

What other response categories are subject to fluctuation as a function of examiner-differences? If examiners studiously attempt to structure three different affectively loaded testing situations, will there be consistent differences, or any differences, as a function of these deliberately varied rapport situations? Will mere repetition of the Rorschach significantly alter any of the scoring categories, regardless of administrator or of affective loading of the administration? If so, which scoring categories are sensitive to such variation?

Thirty-six male college sophomores between the ages of nineteen and twenty-seven participated in a series of three Rorschach tests, administered without testing of the limits. Each test was administered by a different female, and each testing situation was affectively differently loaded. One test was given in the standard manner without pre-tests and without any attempt to give the subject a marked feeling of acceptance or rejection. Another test was administered following two pre-tests during which, by manner and tone of voice, the subject was made to feel rejected and a failure. Still another test was administered after a similar pre-test situation during which the subject was made to feel accepted and successful. Each examiner gave twelve tests first, second and third; twelve tests positively loaded, negatively loaded, and neutral without affective loading. Each subject's different test was administered by a different examiner. An equal number (twelve) of negative, positive, and neutral tests was given on the first, second, and third administrations and by the three different examiners. This plan netted thirty-six Rorschach protocols on each of the three administrations, thirty-six by each of the different examiners, and thirty-six under each of the three different affective loadings of the testing situation. The one hundred eight Rorschach protocols were scored and separately analyzed according to the number of the administration, the administrator, and the type of administration.

On the basis of the analyses of the one hundred eight Rorschach records, the following findings are reported: Subjects who were coldly received tended to misunderstand, and make errors in following directions on simple card-sorting tests. Subjects warmly greeted did not misunderstand, request repetition of, or make errors in following simple directions.

The Experience Balance, or M : sumC ratio, was unstable, only 30% of the subjects maintaining a constant balance throughout the variations in administration. The remainder shifted emphasis from one side of the ratio to the other at least once.

Ten, thirty-one, and forty-eight *t*-ratios were found at the 1 per cent, 5 per cent, and 10 per cent levels of confidence respectively. Of these forty-eight measurable differences, eight were a function of repetition of the test, thirteen a function of the variation in affective loading of the test situation, and twenty-seven a function of examiner differences.

Of the twenty-three Rorschach response-categories subjected to statistical analysis, three were relatively unaffected by the extra-test, or "situational" stimuli: Dd + S%, Fk, and A. The other twenty categories showed frequency differences at least once at the 1 per cent, 5 per cent, or 10 per cent levels of confidence under at least one of the experimentally varied situations.

In conclusion, performance on the Rorschach, as measured by the frequency of responses in the Rorschach categories, varies significantly with repetition of the test, with variation in the negative or positive rapport conditions of the administration, and with examiner differences. In the present study, the largest and most frequent variations in Rorschach performance were associated with examiner differences.

APPENDIX

EXPLANATION OF SCORING SYMBOLS USED IN THIS STUDY[*]

LOCATION

W	Whole blot
W, S	Whole blot and white space used (tabulated as main W and additional S)
D	Large usual detail
D, S	White space used in addition to D (tabulated as main D and additional S)
d	Small usual detail
Dd	Unusual detail (or unusual combinations of usual areas)
S	White space

CONTENT

H	Human figures
Hd	Parts of human figures, not anatomical
A	Animal figures
Ad	Parts of living animals

POPULARITY

P	Popular responses

DETERMINANTS

R	Response
M	Figures in human-like action
FM	Animals in animal-like action
m	Abstract or inaminate movement
k	Shading as three dimensional expanse projected on a two dimensional plane
K	Shading as diffusion
FK	Shading as three dimensional expanse in vista or perspective
F	Form only, not enlivened
Fc	Shading as surface appearance or texture, differentiated
C'	Achromatic surface color
FC	Definite form with bright color
CF	Bright color with indefinite form
C	Color only

$$\text{sum}C = \frac{FC + 2CF + 3C}{2}$$

---

[*] After Klopfer, Bruno and Helen H. Davidson, *The Rorschach Method of Personality Diagnosis.* Individual Record Blank (New York: World Book Company, 1942).

REFERENCES

1. Beck, S. J. *Rorschach's test: I. Basic processes.* New York: Grune & Stratton, 1944.
2. Beck, S. J. *Rorschach's test: II. A variety of personality pictures.* New York: Grune & Stratton, 1945.
3. Bellak, L. The concept of projection. *Psychiatry,* 1944, **7,** 365-366.
4. Bergman, M. S., Graham, H., & Leavitt, H. C. Rorschach exploration of consecutive hypnotic chronological age level regressions. *Psychosom. Med.,* 1947, **9,** 20-28.
5. Bochner, Ruth, & Halpern, Florence. *The clinical application of the Rorschach test.* New York: Grune & Stratton, 1945.
6. Brussel, J. A., & Hitch, K. S. The Rorschach method and its uses in military psychiatry. *Psychiat. Quart.,* 1942, **16,** 3-27.
7. Buhler, Charlotte, Buhler, K., & Lefever, D. W. *Development of the basic Rorschach score.* Los Angeles: mimeographed ed., 1948.
8. Buhler, Charlotte, & Lefever, D. W. A Rorschach study on the psychological characteristics of alcoholics. *Quart. J. Stud. Alcohol,* 1947, **8,** 197-261.
9. Cofer, C. N. Psychological test performance under hyoscine: A case of post-infectious encephalopathy. *J. gen. Psychol.,* 1947, **36,** 221-228.
10. Douglas, Anna Gertrude. A tachistoscopic study of the order of emergence in the process of perception. *Psychol. Monogr.,* 1947, **61,** No. 6.
11. Fosberg, I. A. How do subjects attempt fake results on the Rorschach test? *Rorschach Res. Exch.,* 1943, **7,** 119-121.
12. Fosberg, I. A. Rorschach reactions under varied instructions. *Rorschach Res. Exch.,* 1938, **3,** 12-31.
13. Hall, Calvis S. Diagnosing personality by the analysis of dreams. *J. abnorm. soc. Psychol.,* 1947, **42,** 68-79.
14. Heider, F. Social perception and phenomenal causality. *Psychol. Rev.,* 1944, **51,** 358-374.
15. Hertzman, M., & Pearce, Jane. The personal meaning of the human figure in the Rorschach. *Psychiatry,* 1947, **10,** 413-422.
16. Hsü, E. H. The Rorschach responses and factor analysis. *J. gen. Psychol.,* 1947, **37,** 129-138.
17. Hunt, W. A. The future of diagnostic testing in clinical psychology. *J. clin. Psychol.,* 1946, **2,** 311-317.
18. Jacob, Z. Some suggestions on the use of content symbolism. *Rorschach Res. Exch.,* 1944, **8,** 40-41.
19. Kadinsky, D. Human whole and detail responses in the Rorschach test. *Rorschach Res. Exch.,* 1946, **10,** 140-144.
20. Kamman, G. R. The Rorschach as a therapeutic agent. *Amer. J. Orthopsychiat.,* 1944, **14,** 21-27.
21. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* New York: World Book Co., 1946.
22. Levine, K. N., Grassi, J. R., & Gerson, M. J. Hypnotically induced mood changes in the verbal and graphic Rorschach: A case study. *Rorschach Res. Exch.,* 1943, **7,** 130-144.
23. Levine, K. N., Grassi, J. R., & Gerson, M. J. Hypnotically induced mood changes in the verbal and graphic Rorschach. Part II: The response records. *Rorschach Res. Exch.,* 1944, **8,** 104-124.
24. Lindner, R. M. Content analysis in Rorschach work. *Rorschach Res. Exch.,* 1946, **10,** 121-129.
25. Lindner, R. M. Some significant Rorschach responses. *J. crim. Psychopathol.,* 1944, **5,** 775-778.
26. Lindquist, E. F. Statistical analysis in educational research. Boston: Houghton Mifflin, 1940.
27. Luchins, A. S. Situational and attitudinal influences on Rorschach responses. *Amer. J. Psychiat.,* 1947, **103,** 780-784.
28. Piotrowski, Z. A comparative table of the main Rorschach symbols. *Psychiat. Quart.,* 1942, **16,** 28-37.
29. Rorschach, H. *Psychodiagnostics. A diagnostic test based on perception.* Berne, Switzerland: Hans Huber, 1942. English ed. by Paul Lemkan and Bernard Kronenberg.
30. Schachtel, E. G. Subjective definitions of the Rorschach test situations and their effect on test performance. *Psychiatry,* 1945, **8,** 419-448.
31. Schachtel, E. G. Review, *The Rorschach technique,* by Bruno Klopfer. *Psychiatry,* 1942, **5,** 604-606.
32. U. S. Army Air Forces Aviation Psychology Program Research Report, J. P. Guilford, editor. *Printed classification tests.* Report No. 5. Washington, D. C.: Government Printing Office, 1947. Chapter 24, Clinical type procedures.
33. Wilkins, W. L., & Adams, A. J. The use of the Rorschach test under hypnosis and under sodium amytal in military psychiatry. *J. gen. Psychol.,* 1947, **36,** 131-138.

VOLUME 64
NUMBER 11

WHOLE NO. 317
1950

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

An Evaluation of Personality-Trait Ratings Obtained by Unstructured Assessment Interviews

By

ERNEST C. TUPES

Human Resources Research Center
Lackland Air Force Base
San Antonio, Texas

Based on a dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy
in the University of Michigan
Committee in charge: G. A. Satter, *Chairman;* E. S. Bordin,
E. L. Kelly, W. C. Trow, and E. L. Walker

Accepted for publication, May 8, 1950

Price $1.00

Published by

THE AMERICAN PSYCHOLOGICAL ASSOCIATION
1515 MASSACHUSETTS AVE. N.W., WASHINGTON 5, D.C.

Copyright, 1950 by the
American Psychological Association

ACKNOWLEDGMENTS

THE author wishes to thank his entire committee, and especially Professors George A. Satter and E. Lowell Kelly, for their guidance and constructive criticism in the present study. Additional gratitude is expressed to Dr. Kelly for making available data collected by The Research Project on the Selection of Clinical Psychologists. A special debt of gratitude is owed to the author's wife, Wanda, who by her constant encouragement the past three years, made the present study possible.

ERNEST C. TUPES

TABLE OF CONTENTS

I. Problem, Procedures, and Methods 1
  A. Subjects 1
  B. The Raters 2
  C. The Rating Scale 2
  D. The Interviewers 2
  E. The Initial Interview Situation 3
  F. The Intensive Interview Situation 3
  G. The Design of Assessment 3
    1. The Students' Assessment Schedule 3
    2. Ratings Made during Assessment 4
  H. Data Used in this Investigation 6
  I. The Criterion Measures 6
  J. Method of Determining Validity of Ratings Based on Interviews ... 7
II. Results 9
  A. The Initial Interview 9
    1. Validity of Ratings Made before the Initial Interview 9
    2. Validity of Ratings Made after the Initial Interview 9
    3. The Contribution of the Interview to the Validity of Ratings Made after the Initial Interview 11
  B. The Intensive Interview 12
    1. Validity of Ratings Made before the Intensive Interview 12
    2. Validity of Ratings Made after the Intensive Interview 12
    3. The Contribution of the Interview to the Validity of Ratings Made after the Intensive Interview 13
  C. Comparison of the Intensive and Initial Interviews 13
    1. Agreement between Ratings Made after the Interviews 13
    2. Comparison of the Interviewers' Ratings with respect to Correlation with the *FinP* Ratings 14
      a. The Comparative Validities of Pre-Interview Ratings Based on Different Amounts of Written Material 14

b. The Comparative Validities of Ratings Made after the Initial and Intensive Interview Situations 14
c. The Comparative Validities of Ratings Made before and after the Interviews 15
d. Comparison of Ratings Made after the Initial Interview with Ratings Made before the Intensive Interview 15
3. Comparison of the Interviews with respect to Relative *Net Gains* 15
III. DISCUSSION, CONCLUSIONS, AND SUGGESTIONS FOR FURTHER RESEARCH 17
A. Conclusions 18
B. Suggestions for Further Research 19
IV. SUMMARY 20
A. Objectives of this Study 20
B. Methods and Procedures 20
C. Chief Findings 20
D. Conclusions 21
APPENDIX 22
BIBLIOGRAPHY 24

AN EVALUATION OF PERSONALITY-TRAIT RATINGS OBTAINED BY UNSTRUCTURED ASSESSMENT INTERVIEWS

I. PROBLEM, PROCEDURES, AND METHODS

THE TERM "Interview" as usually defined means almost any face-to-face contact between two or more individuals which involves the exchange of information. When the purpose of the interview is to obtain information about the person being interviewed, it is usually called a diagnostic or appraisal interview. It is with this type of interview that this investigation is concerned.

As pointed out by Fearing (6, 7), the appraisal interview is "probably one of the oldest human social techniques," and is also "the least studied, the most constantly used, and the most frequently challenged method of securing social data."

Studies of the reliability of judgments based on the interview (2, 7, 9, 10, 14, 15), although extremely divergent with respect to method of interview and the type of data selected for analysis demonstrate that these judgments will vary in reliability according to: (*a*) the person doing the interview; (*b*) the amount of objective data available to the interviewer before the interview; and (*c*) the degree of structuring or standardization of the interview itself.

The validity of judgments based on the interview technique has been less adequately studied than has the reliability of interview judgments. In 1931, Symonds (17) pointed out that interviewing "has not been subjected to experimental scrutiny or statistical validation." A survey of the literature by the present writer did not indicate that the situation has been changed since 1931. In none of the studies (1, 4, 5, 10, 13) was any estimate made of the contribution of the interview itself to the validity of the interviewer's judgments. In every case the validities given were for judgments based on the interview plus other material. As a group, the validity studies indicated that little evidence has been gathered to support the belief, prevalent among psychiatrists and psychologists as well as persons in general who deal with other people, that face-to-face contact in an interview situation is very necessary (and sometimes sufficient) for measuring personality traits and predicting behavior.

The present investigation is an attempt to determine the validity of personality-trait ratings based on two types of unstructured interview situations and, further, to determine the increase in validity of interview ratings over that of ratings made by the same individuals without benefit of an interview. The data analyzed in this study were gathered during the 1947 Michigan Assessment Program.[1] If in some instances the data appear incomplete or the results inconclusive it should be understood that the Assessment Program was not designed[2] to thoroughly investigate any particular technique but rather to determine the maximum validity of many techniques combined.

A. SUBJECTS

The sample investigated in this study consisted of 128 male college graduates who had been accepted by various universities for training in clinical psychology leading to the Ph.D. degree under the Veterans Administration training program. All were beginning graduate students accepted at the P-1 level—the lowest professional grade in the U. S. Civil Service system.

---

[1] The 1947 Michigan Assessment Program was conducted as a part of the Research Project on the Selection of Clinical Psychologists sponsored by the Veterans Administration under a contract with the University of Michigan. For a detailed report of the overall aims and research design of this project, the reader is referred to the project's Preliminary Report issued in December, 1948 (11).

[2] The writer, although associated with the project since September, 1946, had no part in the planning or experimental design of the 1947 Michigan Assessment Program and wishes to accept no credit therefor. Neither does he wish to be held responsible for certain justifiable criticisms of the design such as the lack of estimates of the reliability of interview ratings and the lack of independence between interview ratings and criterion ratings.

B. THE RATERS

The subjects were studied intensively over a period of one week by a staff of thirty clinicians. Two of these were psychiatrists. The majority of the others were professional clinical psychologists, some being advanced graduate students in clinical psychology. Ratings made by these clinicians comprise the data of this investigation.

C. THE RATING SCALE

The rating scale used in the Assessment Program comprises 42 variables divided into a Scale A (a group of 22 bi-polar so-called surface traits—supposedly the more manifest dimensions of personality), a Scale B (a group of 9 so-called source traits—supposedly the more underlying dimensions of personality), and a Scale C (a group of 11 so-called criterion variables, which in reality are various skills believed necessary for successful functioning as a clinical psychologist). In addition, ratings were made on *Liking* (Variable #0), defined simply as extent of personal like or dislike for the subject. In rating students on Scales A and B the staff was instructed to rate the person as he was at the time of assessment. In rating on Scale C the student was to be rated as he would be five years in the future (i.e., when he would have one year of professional experience past the Ph.D. degree). Definitions of Scales B and C variables are given in the Appendix.

All three rating scales were the product of joint thinking on the part of the project planning committee and were the outgrowth of trial scales used in preliminary assessments the year before. Scale A was based in part on the findings of Cattell (3) in his factor analyses of personality ratings.

In using the rating scales, raters were instructed to use as a frame of reference "all first year clinical psychology graduate students at universities accredited by the American Psychological Association to offer training in clinical psychology." Raters were asked to make their ratings conform roughly to a normal distribution with suggested frequencies designated as 3, 7, 15, 25, 25, 7, and 3 per cent, respectively, for each point from 1 to 8.

D. THE INTERVIEWERS

A staff team of three members studied each student intensively. This team consisted of an Initial Interviewer, an Intensive Interviewer, and either a Test Integrator or a Projective Integrator. Each staff team studied four students from each class of 24 students. Sixteen staff member functioned as Initial and Intensive Interviewers interchangeably so that for each student team of four students, one staff member would act as Intensive Interviewer for two students and as Initial Interviewer for the other two students. The group of 128 students studied in this investigation thus received two sets of interview ratings—one set made after the Initial interview (hereinafter designated as *Init* ratings) and one set made after the Intensive interview (hereinafter called *Intens* ratings). Since the 16 interviewers acted in both the Initial and Intensive interviewer capacity for the 128 students (although no staff member functioned as both Intensive and as Initial Interviewer for the same student), the problem of interviewer differences does not arise when comparisons are made between ratings based on the Intensive interview situation and ratings based on the Initial interview situation. Each of the interviewers interviewed from two to twelve students in the role of Initial Interviewer and a similar number in the role of Intensive Interviewer.

All 16 interviewers (2 psychiatrists and 14 professional clinical psychologists) had had considerable interviewing experience before assessment. Without doubt they may be regarded as being more skilled at uncovering personality dynamics by use of the unstructured interview than the general population of interviewers. In fact all could undoubtedly be termed expert interviewers.

E. THE INITIAL INTERVIEW SITUATION

The Initial Interview was an unstructured type of interview lasting approximately one hour. The material covered in this interview was purposely kept at a fairly superficial level so that little anxiety would be aroused in the subjects. The Initial Interviewer had available to him before the interview only the information about the subject usually found in a credentials file: i.e., application blank (Civil Service Form 57), letters of recommendation, and records of past performance in the form of college transcripts. The Initial interviewer is probably comparable to the usual college interview for the purpose of determining admission elegibility, and is similar to the typical interview (although somewhat longer in duration) given applicants for employment in industry.

F. THE INTENSIVE INTERVIEW SITUATION

The Intensive interview was also unstructured. It lasted approximately two hours. The Intensive Interviewer had available before the interview the wealth of objective, projective, and autobiographical material described in detail below under *Design of Assessment* (Section I, G, 1). The Intensive interview was deep and probing in nature: the interviewer attempted to uncover underlying personality dynamics in order to confirm hypotheses suggested by his review of the data available on the student, or to fill gaps which persisted in spite of all the data available. The Intensive interview is probably comparable to the interview used for diagnostic purposes with a neuropsychiatric patient by a psychiatrist who has available to him the social history, letters from friends and relatives, and psychological test reports.

G. THE DESIGN OF ASSESSMENT

Lack of space prevents the presenting of the experimental design of the 1947 Michigan Assessment Program in detail.[3] However, some information concerning the assessment is desirable so that the role of the interviews may be seen in proper perspective. Therefore, a typical student ("X") will be followed through the assessment program. In addition, the various assessment ratings will be described in some detail.

1. *The students' assessment schedule.* Students came to the assessment center in groups of 20 to 24 for periods of seven days. Soon after his arrival at assessment headquarters, "X" and his classmates were given an orientation by the Project Director, at which time the aims of assessment were reviewed and the students assured that personal material would be kept in strictest professional confidence. "X" was also told that since certain staff members would be rating him on the basis of certain specified materials he would have absolutely no contact with those staff members until after those particular ratings had been made. The students were then assigned to teams of four.

During his first several days of assessment, "X" took a battery of paper-and-pencil objective tests (Miller Analogies Test, Chicago Test of Primary Mental Abilities [Single Booklet Edition], Allport-Vernon Study of Values, Strong Vocational Interest Blank, Minnesota Multiphasic Personality Inventory, and the Guilford-Martin Battery—Inventory of Factors STDCR, Inventory of Factors GAMIN, and Personnel Inventory I).[4] Scores on these instruments were made available to the assessment staff. The assessee was given a projective test battery consisting of the Rorschach, the Bender-Gestalt, a sentence completion test, and ten cards of the Thematic Apperception Test. He filled out a biographical inventory (consisting of 131 items), a psychological experience record (for aid in evaluating objective and projective tests), and wrote from an outline a detailed autobiography (average length about 20 pages). He was interviewed by an Initial Interviewer, and later interviewed again by an Intensive Interviewer.

---

[3] For a complete description of the assessment see the Preliminary Report (11) referred to earlier.

On the fifth day of assessment, "X" was put through a period of situation or role-playing tests designed to bring out salient features of the student's interactions with others.[5] He was observed and rated at this stage by his own staff team (the two interviewers and the Projective or Test Integrator), and was also observed and rated by another staff team who had never seen him before and who had no information available about him. That evening, all students and staff relaxed with a social gathering.

After breakfast on the last morning, "X" was asked to fill out a sociometric questionnaire (typical questions: "Who would you most like to go on a camping trip with?" and "Which student would you prefer to be supervised by in your work?") concerning his reactions to his classmates. In addition, he was asked to rate himself and his three teammates on Scales A, B, and C (using a slightly modified form of the staff rating form). Finally he was asked to prepare concise but frank character sketches about his teammates.

When these tasks were completed, "X" had a final appointment with his intensive interviewer. Although he hoped to obtain from that interview the staff's opinion of him, he went away satisfied with a small amount of relatively innocuous data, such as test scores.

Finally, "X" and his classmates met for a group-therapeutic session conducted by a visiting psychologist (who had taken no other part in the assessment program) and a farewell talk by the Project Director, who thanked the group for their cooperation and repeated the assurance that assessment findings would be kept in strictest confidence.

2. *Ratings made during assessment.* The preceding section indicates the wide variety of psychological techniques used in assessment. Ideally, the program would have been designed so that each technique could be evaluated separately, with staff members rating independently each student on each technique. Unfortunately, the expense of such a design (in terms of staff and student time as well as money) was prohibitive. As a compromise it was decided to make ratings on various combinations of psychological data and to design the assessment program so that even though estimates of the validity of ratings based on all techniques might not be obtained directly, it would be possible to estimate the increase in validity of ratings obtained with the addition of different psychological data.[6] Wherever practical, in-

---

[4] In addition to these objective tests, the students took the Kuder Preference Record, the ACE General Culture Test, and the Strong Vocational Interest Blank twice more—once with instructions to fill it out as it would be answered by "women in general" and finally as it would be answered by "men in general." Scores on these tests were not made available to the assessment staff but were filed away for later independent analyses.

[5] The situation tests were patterned after the psychodramatic (role-playing) technique of Moreno. For a complete description and thorough analysis of situation tests as used in the 1947 Michigan Assessment Program, the reader is referred to an article by Dr. William Soskin (17). One typical situation test involved two subjects, one of whom was informed in advance that he was to play the role of a small-town high school principal who had called one of his male instructors in for a conference regarding the instructor's rumored sexual misconduct. The other subject was informed that he was to play the role of a high school instructor who had been told the principal wished to talk to him but was not informed what the conference was about. The ensuing five or ten minutes of role-playing consistently called forth aspects of the students' personalities that surprised and enlightened even the participants.

[6] Immediately after each set of ratings was made, the ratings were filed away in the project office and were not again made available to any staff member until the Final Pooling Conference. Similarly, notes made by any rater were not available to any other rater nor were the students discussed with other raters until the Preliminary Conference. Thus, the Intensive Interviewer, for example, when making ratings after the Intensive interview had available only

dependent ratings on various separate psychological techniques were obtained as well. In general, the psychological data were utilized in order of increasing cost; credentials material costing only a postage stamp was presented to the raters first; objective tests costing only slightly more were presented second, etc.

The following ratings were made during assessment:

a. Ratings based on each of the four projective techniques made by the staff member administering the particular technique.

b. Ratings based on the four projective protocols plus interpretations of the protocols. These ratings were made by the Projective Integrators on half the students.

c. Ratings made on the other half of the students by the Test Integrators. These ratings were based on all projective material and in addition on all credentials material, all autobiographical material and all objective test data.

d. Ratings made by the Interviewers.

(1) *PreInit* ratings. These ratings were made by the staff member serving as Initial Interviewer prior to the interview —i.e., before the raters had seen the subjects. These ratings were based entirely on the material contained in the Credentials folders.

(2) *Init* ratings. These ratings were made by the Initial Interviewers immediately after the Initial interviews. Thus, they were based on Credentials materials plus the information obtained during the Initial interview.

(3) *PreIntens* ratings. These ratings were the fourth set of ratings made by the Intensive Interviewers before seeing the subjects. The Intensive Interviewers had previously rated on Credentials material alone; then on Credentials plus Objective test data; then on Credentials plus Objectives plus Autobiographical material. The *PreIntens* ratings were based on Credentials material, Objective test data, Autobiographical material, and Projective protocol material.

(4) *Intens* ratings. These ratings were made by the Intensive Interviewers immediately after the Intensive interview. Thus, they were based on the data described under *PreIntens* ratings above plus the information obtained during the Initial interview.

e. Ratings decided upon at the Preliminary Pooling Conference which occurred after the Intensive Interview. The team of three staff members (the two interviewers and the Projective or Test Integrator) assigned to the student participated in this conference. All the data referred to above were available for this conference as were the notes taken by any of the three staff members prior to the conference. None of the ratings made previously on the student either by the three staff members or by any other staff members were available for the conference.

f. Ratings based on the Situation tests. There were of two types: The three sets of individual Contaminated Situation ratings made by the student's own staff team, and three sets made by another staff team who knew nothing about the student and thus were forced to base their ratings solely upon impressions gained from the Situation tests. In addition, this latter staff team held a Situation Pooling Conference at which they arrived at a set of Uncontaminated Pooled Situation ratings.

---

the written test material and the notes he himself had made during the interview. He had no knowledge of the ratings or the opinions of any other person.

g. The Final Individual ratings. On the last morning of assessment, each member of the student's staff team reviewed all material available on the student (including anything significant he might have gleaned during the social gathering following the Situation tests), and made a set of Final Individual ratings.

h. Ratings by the students. On the last morning of assessment each student rated himself and his three teammates.

i. The *FinP* ratings. During the last afternoon of assessment the student's staff team met for a Final Pooling Conference at which was made available not only all psychological material gathered during assessment, but all ratings that had been made on or by the student. The *FinP* ratings arrived at during this conference thus represent the combined judgment of three clinicians who had intensively studied the subject for seven days and who had available a wealth of psychological data.

H. Assessment Data Used in This Investigation

The *PreInit, Init, PreIntens,*[7] and *Intens* ratings described in the preceding section comprise the ratings whose validities are estimated in this investigation. These ratings were validated against the *FinP* ratings. The estimated reliabilities of the *FinP* ratings are presented in Table 1 (in Section II, *Results*). Unfortunately, the assessment design did not allow for any estimate of the reliabilities of the predictor ratings.

The present paper is concerned only with the validity of ratings on Scales B and C. Scale A was not rated by either of the interviewers prior to the interview, and to present the validities of the interview ratings for Scale A would add nothing to the conclusions drawn from the investigation.

I. The Criterion Measures

Since criterion measures are of paramount importance in any validity investigation (and this is especially true in the area of personality measurement where good criteria are rare indeed), the *FinP* ratings should be examined in detail to determine whether they are acceptable criteria of the personality variables rated.

The estimated coefficients of reliability of the *FinP* ratings (Column 5 of Table 1) on the Scale B variables ranged from .73 to .90 (median .87); those on Scale C from .88 to .93 (median .90). These reliabilities, while not as high as is desirable for individual evaluations, are satisfactory for group measurements. Certainly, they are of such magnitude as to permit evaluation of predictive devices; i.e., failure of any predictor to correlate highly with the *FinP* ratings can only be interpreted as lack of validity of the predictor.

In addition, as noted above, the *FinP* ratings represent the combined and pooled judgments of three staff members, selected on the basis of professional competence in the field of clinical psychology or psychiatry, who had intensively studied each subject for a period of one week, making use of a wide variety of psychological techniques and materials.

Considering these facts, there would seem to be little doubt that the *FinP* ratings of the personality variables making up Scale B are about as valid criterion measures of these variables as are attainable from trained clincians with present techniques.

---

[7] The other three sets of ratings made by the Intensive Interviewers before the interview (see Section G,2,*d*,3 above) are being analyzed elsewhere. Their analyses are not included here because these ratings did not seem especially relevant to the main purposes of this study—the investigation of the validity of ratings based on interviews and increase in validity which can be attributed to the interviews.

The Scale C variables are a different matter. Scale C variables are not personality traits but rather predictions regarding future performance (five years later) in various aspects of the job of Clinical Psychology; Variable C42 was an appraisal of the overall suitability of the student for the field of clinical psychology. It is not believed, therefore, that Scale C *FinP* ratings are acceptable as criterion measures, even though they may represent "better guesses" than ratings made earlier in the assessment program. For this reason comparisons between ratings on Scale C variables based on the interview situations and *FinP* ratings will be considered merely as agreements between two sets of ratings.

One factor which must be considered, if the *FinP* ratings are to be used as criterion measures, is the factor of contamination of the *FinP* ratings. Of the three persons comprising the staff team who made the *FinP* ratings, two were the interviewers whose ratings are here being investigated. If the role of either Intensive Interviewer or Initial Interviewer was regarded consistently by the other two team members as being a "better" role for judging personality characteristics, then that interviewer would have a dominance in the pooling conference which would tend to influence the other two team members. The effect of this influence would be to consistently bring the *FinP* ratings closer to the ratings made by the interviewer than would be warranted by the actual merit of those ratings. Dominance by individuals apart from their roles may be disregarded since all interviewers acted in the dual capacity of Intensive and Initial Interviewers on the same staff team. Staff members were rotated from week to week so that for each student class the staff teams were made up of different combinations of staff members; this also tended to reduce the effect of dominance by any particular individuals.

Detailed evidence on the effect of such role dominance on the estimates of validity of the interviewer's ratings will be presented elsewhere (18). The evidence, while not conclusive, does indicate that role dominance was present; that the role of Intensive Interviewer was believed by the other staff members to be especially advantageous for judging personality variables; but that the effect of such dominance on the *FinP* ratings was slight. The possibility that Intensive Interviewer validity coefficients are spurious will be considered when the results of this investigation are discussed in Section III.

J. METHOD OF DETERMINING VALIDITY OF RATINGS BASED ON INTERVIEWS

To determine the validity of ratings on the various traits made by the interviewers either before or after the interview, the rating given each subject by the interviewer on a particular variable was plotted against the criterion rating given that subject on the same variable. Pearson product-moment correlation coefficients were then computed from the scatter plots, using the conventional raw-score formula. The various sets of correlation coefficients obtained are given in Table 1 appearing in the following chapter.

To determine the contribution of the interview itself to the validity of the ratings made after the interview—that is, the contribution of the interview apart from the material available to the interviewer—correlation coefficients between

TABLE I

SUMMARY OF RESULTS*

Variable	*PreInit-FinP* r	*Init-FinP* r	*PreIntens-FinP* r	*Intens-FinP* r	*FinP* Reliability	*Init-FinP* $r_{x\infty}$	*Intens-FinP* $r_{x\infty}$	*Init Net Gain* Expressed as Per Cent	*Intens Net Gain* Expressed as Per Cent	*Init-Intens* Agreement
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
B 23. Social Adjustment	.02	.38	.50	.61	.90	.40	.64	16	18	.23
B 24. Appropriateness of Emotional Expression	−.07	.20	.49	.62	.85	.32	.67	10	23	.07
B 25. Intensity of Inner Emotional Tension	−.09	.50	.21	.55	.84	.54	.60	30	32	.41
B 26. Sexual Adjustment	−.01	.51	.49	.61	.88	.54	.65	30	20	.36
B 27. Motivation (Status)	.28	.42	.45	.54	.73	.49	.64	15	17	.16
B 28. Motivation (Science)	.27	.28	.52	.63	.88	.30	.67	01	21	.19
B 29. Insight into Others	.22	.32	.36	.50	.86	.34	.54	07	16	.14
B 30. Insight into Self	.21	.44	.48	.63	.87	.47	.68	18	27	.14
B 31. Quality of Intellectual Accomplishment	.40	.56	.70	.77	.90	.59	.81	21	24	.50
Median, Scale B	.21	.42	.49	.61	.87	.47	.65	16	21	.19
C 32. Academic Performance	.54	.65	.76	.78	.91	.68	.82	21	09	.45
C 33. Clinical Diagnosis	.32	.43	.59	.69	.90	.45	.73	10	24	.26
C 34. Individual Psychotherapy	.27	.50	.48	.63	.90	.53	.66	21	25	.24
C 35. Group Psychotherapy	.20	.40	.45	.65	.90	.42	.67	14	39	.31
C 36. Research	.47	.63	.68	.72	.90	.66	.76	26	14	.48
C 37. Administration	.25	.38	.49	.63	.88	.40	.67	10	25	.21
C 38. Supervision	.28	.50	.46	.66	.91	.53	.69	21	33	.29
C 39. Teaching	.39	.49	.68	.74	.90	.52	.78	12	20	.34
C 40. Professional Interpersonal Relations	.16	.37	.43	.62	.91	.39	.65	13	27	.10
C 41. Integrity	.10	.46	.60	.70	.89	.49	.74	23	25	.35
C 42. Overall Suitability	.29	.46	.57	.70	.93	.48	.73	15	28	.37
Median, Scale C	.28	.46	.57	.69	.90	.49	.73	15	25	.31

* $N = 128$. To be significant at the .01 level, r must be .23 or larger; to be significant at the .05 level, r must be .18.

for Scale B and .28 for Scale C. Apparently differences in reliability of the *FinP* ratings are only slightly related to differences in correlations between *Init* ratings and *FinP* ratings.

In order to test the significances of differences between the *Init-FinP* r's for all 20 variables, the r's were first converted into Fisher's *z-functions* by use of a table given by Lindquist (12, p. 212), and the significance of the difference between each pair of z's was estimated. All possible combinations of the 20 variables taken two at a time amount to 189 pairs of differences. Thirty-eight of these differences are significant ($P = .05$ or less). By chance one would expect to find only five differences out of each hundred (or nine out of the 189) reaching the .05 level of significance. The conclusion seems justified that differences do exist in the validities with which the different variables were rated by the *Init* interviewer.[3]

3. *The contribution of the interview to the validity of ratings made after the Initial interview.* Comparison of the *PreInit-FinP* correlations (based on ratings before the interview) with the *Init-FinP* correlations (based on ratings after the interview) provides an opportunity to estimate the incremental contribution of the Initial interview to the validity of the post-interview ratings. If the *PreInit-FinP* r's (Column 1 of Table 1) and the *Init-FinP* r's (Column 2 of Table 1) are squared, the differences between these r^2's provide estimates of the amount of *FinP* rating variance accounted for by the *Init* ratings which can be attributed to the Initial interview. However, the *FinP* ratings differed in reliability and it would seem that these "gain in variance accounted for" figures might be more representative of the contribution of the interview were they corrected in some manner for *FinP* reliability differences. Therefore, an *Init Net Gain* figure was computed by subtracting the *PreInit-FinP* r^2 to obtain a "possible gain in variance accounted for" figure by which the "gain in variance accounted for" figure was divided. It is these relative *Init Net Gains* which are listed in Column 8 of Table 1.[4]

For illustrative purposes, an example of the computation of the *Init Net Gain* figure for Variable B23 follows: From Table 1, it may be seen that for Variable B23, the *PreInit-FinP* r is .02 and the *Init-FinP* r is .38. The difference between the squares of these r's is .1440, which can be interpreted as the amount of the variance of the *FinP* ratings on B23 accounted for by the *Init* ratings which can be attributed to the Initial interview. The reliability of the *FinP* ratings on B23 is .90. Subtracting from this the square of the *PreInit-FinP r* (.0004), the "possible gain in variance accounted for" figure .8996 is obtained. Dividing .1440 by .8996 results in the *Init Net Gain* figure .16, which is given in Column 8 of Table 1.

---

[3] This conclusion is rendered even more tenable by the fact that some correlation might be expected to be present between the z's (since the r's on which the z's are based were all computed on the same sample of cases and hence are not independent), but the amount of this correlation is unknown and thus was ignored in estimating the standard errors of the differences. The effect of ignoring such correlation is to overestimate the standard errors of the differences and thus to underestimate the significance of the differences.

[4] Partial correlation coefficients could have been used to determine the net contribution of the interview to the validity of the post-interview ratings, but when both pre-interview and post-interview validity coefficients are large, a greater difference between the first order r's is necessary to obtain the partial r which would be obtained by a smaller difference between first order r's when their magnitude is smaller. Partial correlation coefficients thus would cause an underestimation of the relative contribution of the interview when the pre-interview ratings correlated higher with the criterion ratings. The *net gains* are not influenced by the magnitudes of the r's, and in addition, by taking account of the reliability of the *FinP* ratings, do not penalize the interviewer's ratings on those variables where, because of lower *FinP* reliabilities, the maximum *Intens-FinP* correlation possible is actually less.

Examination of Column 8 indicates that there are appreciable differences in the *Init Net Gains* between the different variables, but interpretation of these differences is difficult. The two variables with highest *net gains* are B25 and B26, which were among the five variables (18) on which the Final Individual ratings of the Initial Interviewer had the highest (of the three team members) correlations with the *FinP* ratings, so role dominance may be a contributing factor. A common core underlying the variables (B24, B28, B29, C33, and C37) which gained relatively least from the Initial interview becomes apparent when it is considered that these variables all have fairly high loadings on the same factor[5]—one which was tentatively defined as *social intelligence,* or the *ability to effectively use intelligence in interpersonal relations.* Apparently the credentials folder furnished some clues to general intelligence (college transcripts, work history, etc.) which allowed the raters to rate these variables with some success before the interview. The interviewer was able to obtain little information during the Initial interview which added to the estimate of intelligence already made, and little other information which was of help in estimating how effectively the subjects were able to use their intelligence in social situations.

B. THE INTENSIVE INTERVIEW

The mean ratings assigned by the Intensive Interviewers after the interview ranged from 3.89 to 5.54 (median mean rating 4.65). The standard deviations of the *Intens* ratings varied from 1.14 to 1.63 (median standard deviation 1.38).

1. *Validity of ratings made before the Intensive interview.* The correlations between the *PreIntens* ratings (the ratings, made by the Intensive Interviewers just before the interview, based on Credentials, Objective test data, Autobiographical material, and Projective protocols), and the *FinP* ratings appear in Column 3 of Table 1. With one exception (Variable B25) the correlations are significant at or beyond the .01 level of significance. As was found true for the *PreInit* ratings, the variables most validly rated by the Intensive Interviewer before the interview are B31, C32, C36, and C39—all variables which are concerned with intellectual performance.

2. *Validity of ratings made after the Intensive interview.* In Column 4 of Table 1 are listed the validities of ratings made by the Intensive Interviewer after the Intensive interview. In Column 7 appear the validities which would have resulted had the *FinP* ratings been perfectly reliable.

All of the *Intens-FinP* correlations attain significance beyond the .01 level.

It should be noted that correcting the *Intens-FinP r*'s for unreliability of the *FinP* ratings raised the median *r* from .61 to .67 for Scale B, and from .68 to .71 for Scale C. Even had the *FinP* ratings been perfectly reliable the *Intens* ratings would have done only a slightly better job of predicting the *FinP* ratings. The rank difference correlation coefficient, *rho,* computed between the rank orders of the uncorrected *r*'s and the *r*'s corrected for attenuation is equal to .97, indicating that the unreliability of the *FinP* ratings had little effect on the rank order of the validities with which different variables were rated after the Intensive interview. However, differences in the reliability with which the variables were rated by the staff team during the Final Pooling Conference may be related to differences in the validity with which the variables were rated by the Intensive Interviewer. The *rho's* between the rank orders of the *FinP* reliability coefficients and the rank orders of the *Intens-FinP* correlations are equal to .69 for Scale B and .19 for Scale C, suggesting that for Scale B, trait differences in Intensive Interview validities would be smaller were the *FinP* ratings of equal reliability.

---

[5] This factor was obtained in an unpublished factor analysis of the intercorrelations between *FinP* ratings of Scales B and C carried out by the Research Project on the Selection of Clinical Psychologists.

The *Intens-FinP r*'s were converted into *z*'s and tested for significance of differences between them in the manner described for the *Init-FinP r's*. Thirty-eight pairs of differences of the total of 189 pairs were found to be significant at the .05 level, indicating that differences do exist in the correlations between the *Intens* and *FinP* ratings.

3. *The contribution of the interview to the validity of ratings made after the Intensive interview.* The relative *Net Gain* figures for the Intensive interview (obtained by the method described for the Initial interview—Section II, A, 3 above) are shown in Column 9 of Table 1. The variables (B25, B30, C35, C38, C40, and C42) to which the Intensive interview makes the greatest relative contribution seem to be those concerned with *social relations with others in the professional work situation.* The variables to which the Intensive interview contributes very little (B23, B27, B29, C32, and C36) appear to be variables concerned with *effective intelligence* or *intellectual efficiency.* B23 (Social Adjustment) does not fall into this class, but this variable is relatively broad in definition and has many aspects which render it poorly ratable on the basis of a face-to-face conversation. Also, even relatively poorly adjusted individuals might be able to put up a façade for the interview period.

Role dominance may be a contributing factor to the apparent contribution of the Intensive interview. That is, those variables which show relatively high *Net Gains* may do so, not because of any contribution of the interview itself, but because the other two team members, believing the Intensive Interviewer to be the best informed on these variables, deferred to his judgments when determining the *FinP* ratings. In fact, five of the six variables with the highest relative *Intens Net Gains* were variables for which the Final Individual ratings (just before the Final Pooling Conference) made by the Intensive Interviewer correlated higher with the *FinP* ratings than did the Final Individual ratings made by either of the other two staff members (18). Of the five variables with the lowest relative *Intens Net Gains* only two were variables for which the Final Individual ratings of the Intensive Interviewer had higher *r's* with the *FinP* ratings.

C. COMPARISON OF THE INTENSIVE AND INITIAL INTERVIEWS

1. *Agreement between ratings made after the interviews.* In Column 10 of Table 1 are listed the correlations between the *Init* ratings and the *Intens* ratings. Thirteen of the 20 *r*'s were signifi-

cant at the .01 level. Two facts seem of especial interest here. First, the median r for Scale B is .19, while that for Scale C is .31. Apparently, the two interviewers agreed better with each other on ratings of performance five years in the future than they did on ratings of present personality traits. Second, for no variable did the ratings made after the interviews correlate with each other to a greater extent than either set of ratings correlated with the *FinP* ratings. In addition, it may be noted that the variables (B31, C32, and C36) on which the ratings made after the interviews showed the greatest agreement are among those which were most validly rated before the interview by both the Initial and the Intensive Interviewers and which gained least in terms of relative *net gains* from either of the interviews. That is, those variables to the rating of which the interviews contributed relatively little are the very ones on which there is most agreement after the interviews.

2. *Comparison of the Interviewers' ratings with respect to correlation with the FinP ratings.* In Columns 1 through 4 of Table 4 appear four sets of correlations between Interviewers' ratings and *FinP* ratings:

The *PreInit-FinP r*'s (Column 1) are the correlations between the *FinP* ratings and ratings (based on Credentials material only) made by the Initial Interviewer before the interview.

The *Init-FinP r's* (Column 2) are correlations between the *FinP* ratings and ratings by the Initial Interviewer after the one-hour Initial interview.

The *PreIntens-FinP r*'s (Column 3) are correlations between the *FinP* ratings and ratings by the Intensive Interviewer just prior to the interview. In making these ratings, the Intensive Interviewer had available Credentials material, Objective test data, Autobiographical material, and Projective protocols.

The *Intens-FinP r*'s (Column 4) are correlations between the *FinP* ratings and ratings by the Intensive Interviewer after a two-hour Intensive interview.

These four sets of correlations permit of six comparisons, each of which yields some interesting information.

a. The comparative validities of preinterview ratings based on different amounts of written material. When the *PreInit-FinP* correlations are compared with the *PreIntens-FinP*, it is evident that the latter are appreciably higher for all variables. When it is considered that the *PreInit* ratings were based on Credentials material while the *PreIntens* ratings were based on Credentials material plus considerable other psychological data, it seems indicated that the validity of personality-trait ratings will vary with the amount of psychological data made available to the raters.

b. The comparative validities of ratings made after the Initial and Intensive interview situations. The *Intens-FinP* correlations were consistently higher than the *Init-FinP* correlations. Whether the relatively greater validity of ratings after the Intensive interview situation can be attributed to the greater amount of psychological data available to the Intensive Interviewer, or to the greater length and depth of the Intensive interview, or to both factors, is a matter of speculation. The evidence presented below (Section II, C, 3) when the interviews are compared with respect to relative *net gains* suggests that most, if not all, of the apparent superiority of the Intensive interview situation is a result of the greater amount of written data available before the interview.

c. The comparative validities of ratings made before and after the interviews. Four comparisons may be made here. Two of these, the comparison of the validities of ratings before and after the Initial interview and before and after the Intensive interview have been made indirectly in the sections (II, A, 3, and II, B, 3, above) wherein the material concerning the *net gains* in validity attributable to the interview was presented, so it need be only stated here that for all variables, ratings made after each type of interview correlated higher with the *FinP* ratings than did ratings made before that interview. The third comparison, that between *PreInit-FinP* correlations and *Intens-FinP correlations,* merits little consideration since it merely substantiates the rather obvious hypothesis that ratings based on a thorough interview plus a comprehensive mass of written psychological data are more valid than ratings based on only a little written data without benefit of an interview.

The last comparison, that between *Init-FinP* correlations and *PreIntens FinP,* seems of such importance that it is worthy of consideration separately.

d. Comparison of ratings made after the Initial interview with ratings made before the Intensive interview. Proponents of the appraisal interview frequently state that the value of interviews lies in their flexibility. The argument goes that in the interview situation it is possible to follow up leads; to confirm or reject hypotheses regarding the personality dynamics of the interviewee; and thus to form much more accurate judgments than are possible on the basis of test data alone. Test scores, on the other hand, are rigid and inflexible. Even projective devices offer only hints to possible dynamics. Thus, test material may have some value in assessing the more superficial surface traits, but only by talking to the subject, by asking him questions, and by listening to his answers, may the underlying, more fundamental, source traits be accurately appraised.

Comparison of the *Init-FinP* correlations (Column 2 of Table 1) with the *PreIntens-FinP* correlations provides at least a partial test of this belief. The *Init* ratings were based on a one-hour interview (plus the background material available in the Credentials file). The *PreIntens* ratings were based entirely on written data (with no opportunity to talk to, or even see, the subject). The Scale B traits are source traits and hence supposedly more validly rated on the basis of an interview.

This belief is not upheld by the present study. The median *Init-FinP* correlation for Scale B traits is .42; the median *PreIntens-FinP* correlation is .49. The *Init* ratings are more valid for two of the Scale B traits; the *PreIntens* ratings are more valid for seven.

Approximately the same relationship existed for the Scale C variables. The median *Init-FinP* r is .46; the median *PreIntens-FinP* r is .57. The *PreIntens* ratings are more accurate for nine of the eleven variables.

Thus, *ratings based on only test and biographical data are more valid than ratings based on an interview plus the information contained in a credentials folder.*

3. *Comparison of the interviews with respect to relative net gains.* When the relative *net gains* in correlations between ratings which can be attributed to the interviews (Columns 8 and 9 of Table 1) are compared, the Intensive Interviewer seems to gain slightly more from the interview than does the Initial Inter-

viewer, although the difference is neither statistically nor practically significant. The median *net gain* is .24 for the Intensive Interviewer and .16 for the Initial Interviewer. The fact that the Intensive Interviewer appears to have been considered by the other two staff members as being in a somewhat better position to rate many of the variables renders even this small difference questionable.

III. DISCUSSION, CONCLUSIONS, AND SUGGESTIONS FOR FURTHER RESEARCH

THE PRESENT investigation indicates that interviews apparently do contribute to the validity with which personality variables are rated (i.e., the assessment raters typically made more valid ratings on the basis of an interview plus psychological data than they did on the basis of psychological data alone). That contribution is slight, however, and may be even less than has been demonstrated here. It should be remembered that two of the three criterion team members were the interviewers whose ratings were studied. Were the criterion measures truly independent—had none of the criterion team members been interviewers—it does not seem likely that the validity of the interviewers' ratings would have been as high as was found in this study. On the other hand, it is possible that had the interviews occurred first—with psychological data presented to the interviewer after he had already rated on the basis of the interview—their validity coefficients might have been higher, with the psychological test data yielding only a slight additional validity to the interview ratings already made. That is, the psychological data when presented first as in this investigation may have given the interviewers a set so that they were unable to change their ratings after the interview. It seems unlikely that any such persistence of initial impressions actually did decrease the effective validity of the ratings after the interviews for these reasons: (1) The Initial Interviewers, who had a minimum of psychological information (only the credentials files) available before the interview, did not gain more from the interview than did the Intensive Interviewers who had much psychological data available before the interview and thus had opportunity to form stronger initial impressions. (2) Ratings based on unstructured appraisal interviews, unless used in conjunction with some sort of objective psychological data, have been shown to be unreliable, and hence invalid. Additional research is necessary for a definite answer to this question. It is unfortunate that the assessment program could not have been designed to answer this and the similar questions arising when the validity of any of the many assessment techniques is investigated, but, of course, all possible permutations of the various techniques would require several hundred summers of assessments. Since but one assessment was possible it was actually designed on the basis of increasing cost of technique. Thus, credentials costing only postage were made available to the staff first, objective tests second, the interviews (fairly expensive since they required both students and interviewers to be present) later, and last the situation tests requiring groups of students and groups of staff members.

When the two types of interview situations are compared with respect to validity of ratings made by the interviewers after the interviews, it is apparent that the Intensive Interviewer was typically able to make ratings which were more valid than were ratings made by the Initial Interviewer. When the two types of interview situations are compared with respect to relative *net gains* this relationship still exists, but the difference becomes statistically insignificant. In addition, when it is considered that the Intensive Interviewer may have been considered by the other two team members to be the person in the best position to rate, a large portion of this difference in validity would seem to disappear.

There appears, then, to be little difference in the true relative contributions of the two types of interview situations. The Intensive interview, which was two hours in length and in which an attempt was made to probe for underlying personality dynamics, contributed little, if any, more than did the Initial interview, which lasted only one hour and which was ostensibly devoted to the gathering of rather superficial information.

Since differences in the interviews themselves cannot account for the differences in validity of ratings after the two interview situations, these differences would seem to be functions of the differences in the amount of psychological test data available to the interviewers for study before the interview. This is further evidenced by the fact that the *PreIntens* validities were higher than the *PreInit* validities.

Of equal significance is the fact that ratings based on the one-hour interview plus credentials (the *Init* ratings) correlated less well with the criterion ratings than did the *PreIntens* ratings—ratings which were based on a varied and comprehensive set of psychological data but without any interview. Apparently, skilled clinicians can use fairly complete psychological test data in the rating of personality traits slightly more effectively then they can use an interview, even when the interview is aided by historical material in the form of credentials. Unfortunately, neither set of ratings has much effective validity. The median validity of the *Init* ratings for all twenty variables is .45 and the median validity of the *PreIntens* ratings is .49. Such correlation coefficients account for less than 25% of the variance in the criterion ratings, and indicate the predictor ratings to be only about 25% more valid than chance ratings—this despite the fact that the criterion ratings are based on judgments of the same persons who interpreted the psychological data and conducted the interviews. If the Initial interview situation is in fact (as suggested in Section I, E) comparable to the typical personnel selection or college admission interview, it would appear these interviews have little actual value.

Combination of comprehensive psychological data and a two-hour probing interview results in Intensive interview ratings whose median correlation with the criterion ratings is about .63, accounting for about 40% of the criterion variance. This is a sizeable increase over the 20% accounted for by ratings based on psychological data alone, but still leaves the greater percentage of the criterion rating variance unaccounted for—a percentage which would probably be still greater were there not a personnel overlap between the persons making the predictor and the criterion ratings.

Validity coefficients of .65 are considered fairly satisfactory for group tests used for selection purposes, but for individual personality assessment somewhat higher validities are needed. Especially is this true in the areas of clinical psychology and psychiatry where decisions of basic importance to clients and patients must be made—and frequently are made on the basis of much less complete psychological data than were available in the present instance.

A. CONCLUSIONS

The conclusions to be drawn from this investigation have been implicit throughout this and the preceding chapter. Explicitly stated they are:

1. The contribution of the unstructured appraisal interview to the validity of personality-trait ratings based on an interview plus psychological data is slight.

2. Longer interviews with the objective of uncovering personality dynamics contribute little, if any, more to the validity of personality-trait ratings than do shorter interviews whose main objective is the eliciting of information.

3. The more comprehensive the psychological data available, the more valid will be personality-trait ratings based on that data.

4. Personality-trait ratings based on fairly comprehensive psychological test data are more valid than are ratings based on interviews with only a little psychological data in the form of credentials material available.

5. At the present time, even skilled clinicians, having available for study and integration a wide variety of objective test data, projective protocols, credential material, and autobiographical data, plus data accruing from a probing face-to-face interview, do not appear able to make personality assessments, in the form of ratings of personality traits, which have sufficient validity to be as useful in individual case work as is desirable.

B. Suggestions for Further Research

This investigation, although it represents a definite contribution to the study of the validity of personality-trait ratings based on interviews, in that it provides an estimate of the increment in validity which may be attributed to the interview itself, has left at least three important questions unanswered:

1. The previously mentioned problem of whether the presentation of psychological data before the interview formed a "set" in the interviewer which prevented his post-interview ratings being changed sufficiently to reflect the true validity of the interview.

2. The interview as a selection instrument could not be evaluated here since the Final Pooled ratings of Scale C variables did not appear to be acceptable as selection criteria. A similar study is necessary to determine the incremental validity of interview ratings of job-success variables when those ratings are evaluated against acceptable criteria of job-success.

3. Although this investigation was able in some measure to estimate the incremental validity of the interview when used to assess personality variables, it is probable that such incremental validity has been overestimated inasmuch as the criterion measures themselves were determined by the persons who originally made the interview ratings. It would seem desirable, then, that another study be carried out so that the incremental validity of the interview in the assessment of personality-traits might be evaluated against truly independent criteria.

APPENDIX

RATING SCALE DEFINITIONS

SCALE B (Variables Nos. 23-31)

Note: For ratings on this scale, 1 = *left side or low,* 8 = *right side or high.*

Since many of the following attributes (#23-30) are broad factors, it is unlikely that any person will fit all the phrases grouped together at one pole of a given variable. Note also that for some items neither extreme necessarily represents a desirable attribute.

23. *Social Adjustment:* How well does he adjust to varied interpersonal situations? (Includes sexual adjustment only as it affects social adjustment in general.)

Acts without consideration for feelings of others; often rejected by others, often appears aloof, hostile, or irritable.	Actively considers feelings of others; readily gains acceptance in interpersonal relationships; maintains a friendly and likeable manner.

24. *Appropriateness of Emotional Expression:* How appropriate are his emotional responses to the situation?

Fails to adapt his emotional responses to the needs of the situation; shows disorganized or overly constricted emotional responses.	Shows emotional responses of a quality and intensity befitting the situation; reacts spontaneously but appropriately; shows well-integrated and flexible patterns of emotional behavior.

25. *Characteristic Intensity of Inner Emotional Tension:* How intense is his inner emotional life as *inferred* from all available clues?

Inner emotional life characterized by a minimum of persistent internal tensions.	Has strongly repressed emotional drives resulting in inner turmoil; great inner conflict and strong pent-up emotions.

26. *Sexual Adjustment:* To what degree do his sexual needs and activities affect his overall adjustment?

His sexual needs and activities seriously interfere with his overall adjustment.	His sexual needs and activities definitely enhance his overall adjustment.

27. *Motivation for Professional Status:* How strong is his drive for the status-rewards of a professional career?

28. *Motivation for Scientific Understanding of People:* How strong are his drives toward acquiring the facts, theories, and skills necessary for the scientific understanding of individual human beings?

29. *Insight into Others:* How much insight does he have into the attitudes, emotions, and motivations of others?

Interprets behavior at its face value; insensitive to any but gross differences in behavior; does not develop any integrated understanding of behavior or of people.	Has good awareness of underlying dynamics of behavior; is sensitive to subtle nuances of behavioral responses; is able to develop integrated understanding of the behavior of people.

30. *Insight into Himself:* How much insight does he have into the underlying dynamics of his own attitudes, emotions and motivations?

31. *Quality of Intellectual Accomplishments:* What is the characteristic quality of his intellectual output?

Intellectual work is characteristically of low quality.	Characteristically produces intellectual work of high quality.

SCALE C—CRITERION SKILLS (Variables Nos. 32-42)

Note: For ratings on this scale, 1 = *low,* 8 = *high.*

Ratings on No. 32 refer to performance in graduate school; ratings on Nos. 33-42 refer to student's performance five years hence (i.e., after one year of experience past the Ph.D.).

What will be his level of competence or skill in the varied aspects of:

32. *Academic Performance* (during next three or four years): How well will he:

Effectively master course content, successfully complete courses in general psychology, clinical psychology, statistics, and related fields; satisfy language requirements for the doctorate; pass general examinations.

33. *Clinical Diagnosis:* How well will he:

Recognize dynamics underlying particular responses in both objective and projective tests, observe significant interrelationships among responses, relate findings to case history and other clinical data. Elicit from the patient information required for mental status examinations and case histories; ascertain and evaluate attitudes and incidents of psychological significance in the patient. Synthesize clinical findings to arrive at an integrated picture of personality development, structure, and function.

34. *Individual Psychotherapy:* How effectively will he: Conduct various types of individual psychotherapy.

35. *Group Psychotherapy:* How effectively will he: Conduct various types of group psychotherapy.

36. *Research:* How well will he:

Recognize and define important research problems in clinical psychology; critically evaluate and apply the research findings of others; think with originality and scientific rigor; employ appropriate experimental design and statistical methods; grasp practical implications of findings; present results and conclusions in clear, comprehensive, and well-organized form.

37. *Administration:* How well will he:

Plan and develop psychological programs; make proper administrative decisions; delegate responsibility appropriately; elicit cooperation from subordinates and superiors; maintain high morale among his staff; carry out or direct an appropriate public relations program.

38. *Supervising Clinical Psychologists:* How well will he:

Carry out the professional supervision of subordinates assigned to him for duty and on-the-job instruction; assign their duties; evaluate their performance; instruct them in clinical techniques; perform other aspects of in-service training.

39. *Teaching Psychology* (in a College or University): How well will he:

Teach college courses in general psychology; motivate students; present concepts and procedures; stimulate critical thinking about and integration of course materials; evaluate the products of learning.

40. *Professional Interpersonal Relations:* How well will he:

Work cooperatively with superiors, subordinates, members of the mental team, and other professional personnel concerned with the patient's welfare; participate in the give-and-take of staff conferences; contribute to group decisions.

41. *Integrity of Personal and Professional Behavior:* How well will he:

Recognize and fulfill professional responsibilities; live up to personal commitments; show loyalty to professional obligations in the event of outside pressure or promise of personal gain; maintain discretion concerning professional matters; appropriately conform with commonly accepted standards of moral and social behavior; refrain from coloring facts, evasion, lying, etc.

42. *Overall Suitability for Clinical Psychology:* In view of his assets and liabilities, how well will he be able to:

Carry out the several duties—diagnosis, therapy, and research—specified for the position of clinical psychologist (P-4 and above) in the Veterans Administration.

BIBLIOGRAPHY

1. *Assessment of men.* OSS Assessment Staff. New York: Rinehart, 1948.
2. Bingham, M. V. D., and Moore, B. V. *How to interview.* New York: Harper, 1941.
3. Cattell, R. B. *Description and measurement of personality.* Yonkers, N.Y.: World Book Co., 1946.
4. Clark, E. B. Value of student interviews. *J. Person. Res.*, 1926, **5**, 204-207.
5. Escalona, Sybille. *Progress report: research project on the selection of medical men for psychiatric training.* Topeka: Menninger Foundation, 1948. (Mimeographed and privately circulated.)
6. Fearing, Franklin. The appraisal interview: a critical consideration of its theory and practice with particular reference to the selection of public personnel. In *Studies in personality.* New York: McGraw-Hill, 1942.
7. Fearing, F., and Fearing, F. M. Factors in the appraisal interview considered with particular reference to the selection of public personnel. *J. Psychol.*, 1942, **14**, 131-153.
8. Fiske, Donald W. Consistency of the factorial structures of personality ratings from different sources. *J. abnorm. soc. Psychol.*, 1949, **44**, 329-344.
9. Hollingworth, H. L. *Vocational psychology and character analysis.* New York: Appleton, 1929.
10. Hovland, C. I., and Wonderlic, E. F. Prediction of industrial success from a standardized interview, *J. appl. Psychol.*, 1939, **26**, 537-546.
11. Kelly, E. L., and Fiske, D. W. *The selection of clinical psychologists: progress report and preliminary findings of the research project on the selection of clinical psychologists.* Ann Arbor, Michigan, 1948. (Lithographed and privately circulated.)
12. Lindquist, E F. *Statistical analysis in educational research.* Boston: Houghton Mifflin, 1940.
13. Moss, F. A. Scholastic aptitude tests for medical students, *J. Ass. Amer. Med. Colleges*, 1931, **6**, 1-16.
14. Newman, S. H., Bobbitt, J. M., and Cameron, D. C. The reliability of the interview method in an officer candidate evaluation program. *Amer. Psychologist*, 1946. **1**, 103-110.
15. Scott, W. D., Bingham, M. V., and Whipple, G. M. The scientific selection of salesmen. *Salesmanship*, 1916, **4**, 106-108.
16. Soskin, W. F. Standardized situations as a means of personality appraisal. *Amer. Psychologist*, 1948, **3**, 271.
17. Symonds, P. M. *Diagnosing personality and conduct.* New York: Appleton-Century, 1931.
18. Tupes, E. C. *The effect of role dominance on estimates of validity of interviewers' ratings* (in preparation).
19. Tupes, E. C. *Changes in means and standard deviations of personality-trait ratings after interviews* (in preparation).

0
o